Quality and Reliability Engineering International

WILEY

RESEARCH ARTICLE

Estimating the Generalizations of Process Capability Index C_{pmk} in the Presence of Outliers

Hamideh Iranmanesh[1] | Mehdi Jabbari Nooghabi[1] | Abbas Parchami[2]

[1]Faculty of Mathematical Sciences, Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran | [2]Faculty of Mathematics and Computer, Department of Statistics, Shahid Bahonar University of Kerman, Kerman, Iran

Correspondence: M. Jabbari Nooghabi (jabbarinm@um.ac.ir)

Received: 8 July 2024 | **Revised:** 18 November 2024 | **Accepted:** 23 November 2024

Funding: The authors received no specific funding for this work.

Keywords: maximum-likelihood estimator | moment estimator | outliers | robust estimator

ABSTRACT

Process capability indices (PCIs) play a fundamental role in assessing and quantifying the capability of manufacturing processes to meet customer specifications. To date, the majority of PCIs that are now in use have been examined through the use of novel generalizations of PCIs with symmetric tolerances to check the performance of an industrial manufacturing process. In this paper, we proposed new estimations of some indices with asymmetric tolerances in the presence of outliers. New robust and parametric estimators of some PCIs are introduced to estimate and compare these indices for each normal distribution with the presence of different outliers. Meanwhile, this paper discusses how well the proposed method can be used for non-normal data. For illustration purpose, the application example is presented.

1 | Introduction

Over three recent decades, process capability indices (PCIs) have become an essential tool for process improvement and quality management in various industries. They have been widely employed to assess the capability of manufacturing processes by providing numerical metrics that specify whether a process complies with the capability requirements set in industrial manufacturing factories. Statistical process control (SPC) has made process capability analysis (PCA) a vital component that is used to improve the quality. Therefore, the manufacturing department can enhance the process to raise the quality level and meet customer requirements by analyzing PCIs. A well review regarding the process capability analysis can be found in Kotz and Johnson [1]. The four widely recognized PCIs have been presented in the literatures as follows:

$$C_p = \frac{USL - LSL}{6\sigma},$$

$$C_{pk} = \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\},$$

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}}$$

and

$$C_{pmk} = \min\left\{\frac{USL-\mu}{3\sqrt{\sigma^2+(\mu-T)^2}}, \frac{\mu-LSL}{3\sqrt{\sigma^2+(\mu-T)^2}}\right\} = \frac{d - |\mu - m|}{3\sqrt{\sigma^2+(\mu-T)^2}},$$

in which μ is the process mean, σ is the process standard deviation, USL and LSL are the upper and lower specification limits, T is the target value, $m = (USL + LSL)/2$ is the midpoint of the specification limits, and $d = (USL - LSL)/2$ is half of the length of the specification interval; see Kane [3]. The PCI C_p measures the process variation related to the preset SLs. Moreover, Boyles [4] proposed the PCI

$$S_{pk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\mu}{\sigma}\right)+\frac{1}{2}\Phi\left(\frac{\mu-LSL}{\sigma}\right)\right\}, \quad (1)$$

for a normal process, in which Φ is the cumulative distribution function of the standard normal distribution and Φ^{-1} represents its inverse function. These proposed PCIs aim to monitor only the performance in processes with symmetric tolerances (that is, $T = m$), based on normal, independent, and statistically controlled observations. It is essential to highlight that a higher index value indicates a more capable process. Although the common perception is that symmetric cases are predominant, it is important to acknowledge that situations often arise where the target value is not located at the midpoint of the tolerance (i.e., $T \neq m$). This condition, known as asymmetric tolerance, is a common occurrence in industrial production factories. Examples of asymmetric tolerances occur in specific situations. These differences are usually not exclusively related to the form of the supplier's process distribution. Rather, they indicate that certain directions of departure from the target are more acceptable than others; see Vännman [5]. Because of the various quality characteristics of products, practitioners are not restricted by the standard specification setting. Moreover, many initially asymmetric tolerances stem from the initial condition of symmetric tolerances, but as time progresses, the process adheres to a distribution that is not normal. This is a rather common way that asymmetric tolerances develop. Specifically, transforming data to approximate normality can give rise to asymmetric tolerances. Following this transformation, it becomes feasible to convert symmetric tolerances into asymmetric ones.

There has been relatively little attention of the asymmetry of the specification limits. Boyles [4] recommended the PCI S_{pmk} as an extension of PCI C_{pmk}. A generalization of C_{pmk}, known as C''_{pmk}, was also presented by Pearn et al. [6] to cover processes with asymmetric tolerances. In this paper, we proposed the estimations of indices S_{pmk} and C''_{pmk} in the presence of outliers. Numerous research papers have focused on addressing processes with asymmetric specification limits for the C_{pmk} index and two alternative extensions of C_{pmk}, including

$$S_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\mu}{\sqrt{\sigma^2+(\mu-T)^2}}\right) + \frac{1}{2}\Phi\left(\frac{\mu-LSL}{\sqrt{\sigma^2+(\mu-T)^2}}\right)\right\} \quad (2)$$

and

$$C''_{pmk} = \frac{d^*-A^*}{3\sqrt{\sigma^2+A^2}}, \quad (3)$$

where $A = \max\{d(\mu-T)/D_u, d(T-\mu)/D_l\}$, $A^* = \max\{d^*(\mu-T)/D_u, d^*(T-\mu)/D_l\}$, $D_l = T - LSL$, $D_u = USL - T$, and $d^* = \min\{D_l, D_u\}$. If $T = m$, therefore the index C''_{pmk} reduces to the index C_{pmk}. Nevertheless, it might underestimate or overestimate process capability in various instances, contingent upon the relationship between μ and T. The findings indicated that C''_{pmk} provides a more precise assessment of process capability compared to the index C_{pmk} and other current generalizations of C_{pmk} for processes with asymmetric tolerances; see Boyles [4] and Pearn et al. [6, 7].

Broadly, employing PCIs is essential to measure the extent to which process outputs meet the predetermined capability standards. The presence of outliers can obscure the identifiable sources of variation, potentially leading to unreliable results when utilizing PCIs. Moreover, μ and σ are unknown parameters in the PCIs, and therefore we require a random sample to estimate the unknown parameters. Hence, μ and σ must be estimated for estimating the PCI. In many studies, such as Iranmanesh et al. [8–10] and Parchami et al. [11, 12], it is common to use the natural estimator of the considered PCI. It should be noted that the natural estimator of the considered PCI is created by substituting $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $S_{n-1} = \sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/n-1}$ instead of the unknown parameters μ and σ in the considered PCI, respectively. Moreover, the existence of outliers can significantly impact on statistical analyses. Lu and Chang [13] proposed a robust procedure for solving multiphase regression problems that is efficient enough to deal with data contaminated by atypical observations due to measurement errors or those drawn from heavy-tailed distributions. Performances of two robust PCIs for multiple linear profiles in comparison with the classical PCIs in the absence and presence of contamination were evaluated in [14]. A few robust estimators for the capability index S_{pk} were defined by Iranmaneshet al. [15]. In Prasad and Bramorski [16], resilient time series methodologies were explored to establish novel collections of PCIs applicable to a diverse array of industrial processes. The motivation for estimating the PCIs on the basis of the parametric model of outliers was discussed in Jabbari Nooghabi [17].

The intention of this paper is presenting new estimators for the generalizations of C_{pmk} with asymmetric tolerances to evaluate the performance of the manufacturing process in the presence of outliers. For this intention, new parametric and robust estimators of the indices S_{pmk} and C''_{pmk} are introduced to estimate and compare these indices with the presence of different outliers. Hence, these indices are estimated on the basis of the robust, maximum-likelihood (ML), and method of moment (MM) estimators of the unknown parameters of the normal distribution contaminated by outliers. It has been observed that parametric estimations have better performances than the robust estimations. This paper is organized in the following structure. Section 2 contains the preliminaries and presents some new estimations of PCIs in the presence of outliers. Section 3 incorporates a comparison study between the parametric and robust estimation procedures. Also, this section discusses how well the proposed procedures can be used for non-normal data. The illustrative results are detailed in Section 4. Lastly, conclusions and future works are provided.

2 | New Estimations of the PCIs in the Presence of Outliers

In the last 30 years, the term "outlier" has been a topic of continuous discussion in academic literature. An observation in a dataset that substantially differs from the remaining recorded data points is called an outlier; see Jabbari Nooghabi [17]. When a dataset contains one outlier or just a few of them, we encounter a substantial challenge in estimating parameters. In this context, parametric/robust estimation procedures can be highly suitable for measuring the capability of the process.

There has been extensive discussion about how the existence of outliers can negatively impact on statistical analyses and decision-making processes; see more details in Dixit [18], Dixit and Jabbari Nooghabi [19, 20], and Jabbari Nooghabi [17]. In this regard, we proposed new parametric and robust estimators of the indices S_{pmk} and C''_{pmk} to estimate and compare these indices in the presence of different outliers. Herein, by inspiration of Iranmanesh et al. [15], we define new robust estimators of the indices S_{pmk} and C''_{pmk} to measure the performance the manufacturing process for processes with asymmetric tolerances for every normal distribution. Additionally, we apply the proposed parametric method by Jabbari Nooghabi [17] to estimate the proposed indices as more useful estimation method than the robust estimation methods in the presence of outliers. In this section, we intend to introduce some new estimations of process capability indices with the presence of outliers.

2.1 | Estimation Based on Robust Procedures

Typically, two straightforward robust estimators for the scale and location parameters are the median absolute deviation (MAD) and the median, respectively. The MAD serves as one of the substitutes to the robust estimation of the standard deviation. Consequently, a definition of the MAD is provided by Hampel [21]:

$$\text{MAD} = \underset{i}{\text{median}}|X_i - M|, \tag{4}$$

in which M represents the sample median, serving as a robust estimator for μ. Herein, 1.4826 MAD, often referred to the standardized MAD, is employed as a consistent robust estimator of σ under the normal distribution; see Rousseeuw and Croux [22]. Additionally, the less robust but simpler estimate for the scale parameter is the interquartile range (IQR).

Definition 2.1. Consider IQR to be the difference between the third quartile and the first quartile of the sample data ($\text{IQR} = Q_3 - Q_1$). Subsequently, a threshold based on IQR is defined as follows:

$$\begin{cases} T_{min} = Q_1 - 1.5\,\text{IQR}, \\ T_{max} = Q_3 + 1.5\,\text{IQR}, \end{cases} \tag{5}$$

where T_{min} and T_{max} represent the minimum and maximum thresholds for identifying outliers. Typically, a data point falling outside the interval $[T_{min}, T_{max}]$ is classified as an outlier in Yang et al. [23].

Following the ideas presented in Iranmanesh et al. [15], we introduce two new robust estimators for the indices S_{pmk} and C''_{pmk} to estimate them effectively, particularly in the presence of outliers. The IQR- and MAD-based estimators of S_{pmk} to handle the processes with asymmetric tolerances are introduced as follows:

$$\hat{S}^{\text{IQR}}_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL - M}{\tau_{\text{IQR}}}\right) + \frac{1}{2}\Phi\left(\frac{M - LSL}{\tau_{\text{IQR}}}\right)\right\} \tag{6}$$

and

$$\hat{S}^{\text{MAD}}_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL - M}{\tau_{\text{MAD}}}\right) + \frac{1}{2}\Phi\left(\frac{M - LSL}{\tau_{\text{MAD}}}\right)\right\}, \tag{7}$$

where $\tau_{\text{IQR}} = \sqrt{(\text{IQR}/3)^2 + (M - T)^2}$, $\tau_{\text{MAD}} = \sqrt{(1.4826\ \text{MAD})^2 + (M - T)^2}$, in which T, IQR, M, and 1.4826 MAD represent the target value, the sample IQR, the sample median, and the sample standardized MAD, respectively. Also, the MAD- and IQR-based estimators of C''_{pmk} to assess the performance of the processes with asymmetric tolerances are defined as follows:

$$\hat{C}''^{\text{MAD}}_{pmk} = \frac{d^* - A^*_M}{3\sqrt{(1.4826\ \text{MAD})^2 + A_M^2}}, \tag{8}$$

and

$$\hat{C}''^{\text{IQR}}_{pmk} = \frac{d^* - A^*_M}{3\sqrt{(\text{IQR}/3)^2 + A_M^2}}, \tag{9}$$

where $A_M = \max\{d(M - T)/D_u, d(T - M)/D_l\}$, $D_l = T - LSL$, $D_u = USL - T$, $d^* = \min\{D_l, D_u\}$, and $A^*_M = \max\{d^*(M - T)/D_u, d^*(T - M)/D_l\}$.

2.2 | Estimation Based on Parametric Procedures

By inspiration of Jabbari Nooghabi [17], we are going to a presentation of the parametric procedures with the presence of outliers generated from the normal distribution. Outliers could originate from the identical distribution with different parameters. The normal distribution is characterized by the following two parameters: the mean and standard deviation. Hence, we intend to consider that the distribution of outliers differs specifically in terms of the mean, indicating a shift in the mean. Suppose that there are n observed data, where k data have a normal distribution with parameters $\mu + \delta$ and σ^2 and the remaining $n - k$ out of n observed data follow a normal distribution as $N(\mu, \sigma^2)$. Hence, applying the general probability rule, the probability density function (pdf) of the normal distribution with the presence of k outliers for one observation x can be expressed as follows; see more details in Jabbari Nooghabi [17]:

$$f(x; \mu, \delta, \sigma^2) = bN(\mu + \delta, \sigma^2) + \bar{b}N(\mu, \sigma^2),\ x, \mu, \delta \in \mathbb{R},\ \sigma^2 > 0,\ 0 < b, \bar{b} < 1, \tag{10}$$

where $b = k/n$ and $\bar{b} = 1 - k/n$. Therefore, the joint density of $X_1, X_2, \ldots, X_n$ is as follows:

$$f(x_1, x_2, \ldots, x_n; \mu, \delta, \sigma^2) = \frac{(2\pi\sigma^2)^{-n/2}}{C(n,k)} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$
$$\exp\left\{ \frac{-k\delta}{\sigma^2} \left(\frac{\delta}{2} + \mu \right) \right\} \sum_{i_1, i_2, \ldots, i_k} \prod_{j=1}^{k} \exp\left\{ \frac{\delta}{\sigma^2} x_{i_j} \right\}. \tag{11}$$

in which $C(n,k) = \frac{n!}{k!\,(n-k)!}$ and $\sum_{i_1, i_2, \ldots, i_k} = \sum_{i_1=1}^{n-k+1} \sum_{i_2=i_1+1}^{n-k+2} \cdots \sum_{i_k=i_{k-1}+1}^{n}$.

Notation 2.2. Recognize that the number of outliers, k, is a known parameter in the model as well as the sample size. However, in practical terms, the number of outliers is unknown and determining the exact number of outliers is challenging. Consequently, when faced with this uncertainty, the value of k can be determined by assessing the likelihood function across various k values and selecting the one that maximizes the likelihood function.

Herein, by adapting from Jabbari Nooghabi [17], we intend to estimate the parameters of the normal distribution in the presence of outliers. These parameters are estimated with respect to the mean shifted, and are taken into consideration in the following points on the basis of the MM and ML procedure. Define all the parameters of the normal distribution contaminated with outliers are unknown. Then, for the instance where the mean was shifted, three parameters μ, δ, and σ are unknown.

1. **MM procedure:** Based on considering the first, second, and third moments of the normal distribution in the presence of outliers, one can estimate the unknown parameters [17]. Therefore, the sample moments are shown by $\bar{x^a} = \sum_{j=1}^{n} x_j^a / n$, for $a = 1, 2, 3$, and according to the MM, we have the following equations:

$$\begin{cases} \hat{\delta}_{MM} = \sqrt[3]{\dfrac{2\bar{x}^3 - 3\bar{x}\bar{x^2} + \bar{x^3}}{2b^3 - 3b^2 + b}}, \\ \hat{\mu}_{MM} = \bar{x} - b\hat{\delta}_{MM}, \\ \hat{\sigma}^2_{MM} = -\bar{x}^2 + b^2\hat{\delta}^2_{MM} - b\hat{\delta}^2_{MM} + \bar{x^2}. \end{cases} \tag{12}$$

2. **ML procedure:** From Equation (11), the log-likelihood function for the observed values $x_1, \ldots, x_n$ is as the following form:

$$l(\mu, \delta, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \ln(C(n,k)) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$
$$\frac{-k\delta}{\sigma^2} \left(\frac{\delta}{2} + \mu \right) + \ln\left(\sum_{i_1, i_2, \ldots, i_k} \prod_{j=1}^{k} \exp\left\{ \frac{\delta}{\sigma^2} x_{i_j} \right\} \right). \tag{13}$$

Hence, by inspiration of Jabbari Nooghabi [17], the ML estimators for the parameters are acquired through differentiation with respect to them. Consequently, the ML estimators for the parameters μ, δ, and σ, denoted as $\hat{\mu}_{ML}$, $\hat{\delta}_{ML}$, and $\hat{\sigma}_{ML}$, respectively, are determined by solving the following three equations:

$$\frac{1}{\hat{\sigma}^2_{ML}} \sum_{i=1}^{n} (x_i - \hat{\mu}_{ML}) - \frac{k\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} = 0, \tag{14}$$

$$-\frac{k\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} - \frac{k\hat{\mu}_{ML}}{\hat{\sigma}^2_{ML}} + \frac{\sum_{i_1, i_2, \ldots, i_k} \sum_{j=1}^{k} x_{i_j} \exp\left\{ \frac{\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} x_{i_j} \right\}}{\hat{\sigma}^2_{ML} \sum_{i_1, i_2, \ldots, i_k} \exp\left\{ \frac{\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} \sum_{j=1}^{k} x_{i_j} \right\}} = 0 \tag{15}$$

and

$$-\frac{n\hat{\sigma}^2_{ML}}{2} + \frac{1}{2} \sum_{i=1}^{n} (x_i - \hat{\mu}_{ML})^2 + \frac{k\hat{\delta}^2_{ML}}{2} + k\hat{\delta}_{ML}\hat{\mu}_{ML}$$
$$-\frac{\hat{\delta}_{ML} \sum_{i_1, i_2, \ldots, i_k} \sum_{j=1}^{k} x_{i_j} \exp\left\{ \frac{\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} x_{i_j} \right\}}{\sum_{i_1, i_2, \ldots, i_k} \exp\left\{ \frac{\hat{\delta}_{ML}}{\hat{\sigma}^2_{ML}} \sum_{j=1}^{k} x_{i_j} \right\}} = 0, \tag{16}$$

Herein, employing mathematical techniques does not allow for the derivation of a closed form solution. Therefore, the ML estimators of the parameters μ, δ, and σ (i.e., $\hat{\mu}_{ML}$, $\hat{\delta}_{ML}$, and $\hat{\sigma}_{ML}$, respectively) are obtained by using numerical methods.

Notation 2.3. In maximum likelihood estimation, the likelihood function is derived from the probability model representing the data, and numerical optimization methods are employed to find the parameter values that maximize this likelihood function. The process of parameter initialization and its effect on convergence and final estimates is critical, especially when using iterative numerical methods. The initialization of parameters in MLE plays a crucial role in the success of numerical optimization methods. It should be noted that we use the multiroot function, available in the "rootSolve" package in R [24], to solve Equations (14)–(16). To address this, we present the following step-by-step guide on how initial values are chosen for numerical methods:

1. First, the initial guess for the parameter μ used as the starting value for the multiroot function is determined as the sample mean, $\mu^* = \bar{x}$.
2. In the second step, consider using the initial guess for the parameter δ as the starting value for the multiroot function, based on the following formula:

$$\delta^* = \max\left\{ |\bar{x} - \max_{i=1,\ldots,n} x_i|, |\bar{x} - \min_{i=1,\ldots,n} x_i| \right\}. \tag{17}$$

3. In the third step, the initial guess for the parameter σ used as the starting value for the multiroot function is considered to be the sample standard deviation, $\sigma^* = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)}$.
4. Finally, the initial values (μ^*, δ^*, and σ^*) are used to solve Equations (14)–(16) to obtain the ML estimators of the parameters μ, δ, and σ (i.e., $\hat{\mu}_{ML}$, $\hat{\delta}_{ML}$, and $\hat{\sigma}_{ML}$, respectively).

In the presence of outliers, the parametric model of outliers is taken into consideration to define the parametric estimators of

S_{pmk} and C''_{pmk}, which are more useful than the presented robust estimators; see Jabbari Nooghabi [17].

Definition 2.4. Let $\hat{\mu}_{MM}$, $\hat{\sigma}_{MM}$, $\hat{\mu}_{ML}$, and $\hat{\sigma}_{ML}$ be the method of moment estimators (MMEs) and the maximum-likelihood estimators (MLEs) of the unknown parameters μ and σ, respectively. Then, the parametric estimators, denoted as the ML- and MM-based estimators of S_{pmk}, are introduced as follows:

$$\hat{S}^{MM}_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\hat{\mu}_{MM}}{\sqrt{\hat{\sigma}^2_{MM}+(\hat{\mu}_{MM}-T)^2}}\right) + \frac{1}{2}\Phi\left(\frac{\hat{\mu}_{MM}-LSL}{\sqrt{\hat{\sigma}^2_{MM}+(\hat{\mu}_{MM}-T)^2}}\right)\right\}, \tag{18}$$

and

$$\hat{S}^{ML}_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL-\hat{\mu}_{ML}}{\sqrt{\hat{\sigma}^2_{ML}+(\hat{\mu}_{ML}-T)^2}}\right) + \frac{1}{2}\Phi\left(\frac{\hat{\mu}_{ML}-LSL}{\sqrt{\hat{\sigma}^2_{ML}+(\hat{\mu}_{ML}-T)^2}}\right)\right\}. \tag{19}$$

Definition 2.5. Let $\hat{\mu}_{MM}$, $\hat{\sigma}_{MM}$, $\hat{\mu}_{ML}$, and $\hat{\sigma}_{ML}$ be the MMEs and MLEs of the unknown parameters μ and σ, respectively. Then, the ML- and MM-based estimators of C''_{pmk} are introduced as follows:

$$\hat{C}''^{MM}_{pmk} = \frac{d^*-\hat{A}^*_{MM}}{3\sqrt{\hat{\sigma}^2_{MM}+\hat{A}^2_{MM}}}, \tag{20}$$

and

$$\hat{C}''^{ML}_{pmk} = \frac{d^*-\hat{A}^*_{ML}}{3\sqrt{\hat{\sigma}^2_{ML}+\hat{A}^2_{ML}}}, \tag{21}$$

where

$$\hat{A}_{MM} = \max\{d(\hat{\mu}_{MM}-T)/D_u, d(T-\hat{\mu}_{MM})/D_l\},$$
$$\hat{A}^*_{MM} = \max\{d^*(\hat{\mu}_{MM}-T)/D_u, d^*(T-\hat{\mu}_{MM})/D_l\},$$
$$\hat{A}_{ML} = \max\{d(\hat{\mu}_{ML}-T)/D_u, d(T-\hat{\mu}_{ML})/D_l\},$$
$$\hat{A}^*_{ML} = \max\{d^*(\hat{\mu}_{ML}-T)/D_u, d^*(T-\hat{\mu}_{ML})/D_l\},$$
$$D_l = T-LSL, D_u = USL-T \text{ and } d^* = \min\{D_l, D_u\}.$$

3 | A Comparison Study Between Parametric and Robust Estimation Procedures

We intend to present a comparative analysis of the mean square error (MSE) criterion on the basis of the IQR-, MAD-, MM-, and ML-based estimators in this section. In order to evaluate how the robustness of various estimation methods is to outliers, we take into consideration m random samples, each with n observations taken from a normal distribution. These samples encompass different numbers of outliers denoted by k. The MSE of the parametric and robust estimators of S_{pmk} and C''_{pmk} for various numbers of outliers are provided in Tables 1 and 2. Also, Figures 1 and 2 are the representations of the intuitive understanding of Tables 1 and 2. It must be noted that the MSE by utilizing m replications, is computed as follows:

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i-\theta)^2, \tag{22}$$

where $\hat{\theta}_i$ is the estimator of θ for $i = 1, \dots, m$.

Notation 3.1. In a comparative study between the proposed parametric and robust estimation procedures, the MSE criterion is used to demonstrate that the proposed parametric estimators are more suitable for estimation tasks involving the presence of outliers. The lower MSE value of the proposed estimator indicates that it performs better than other estimators. To clarify the computational complexity, we present the following step-by-step guide on how the MSE values are calculated for the proposed parametric and robust estimators of S_{pmk} and C''_{pmk} (see more details about this computation in Algorithm 1). Herein, we suppose that the outliers follow the same distribution with different parameters in this investigation. For a sample with size n in the presence of k outliers, we generate $m = 1000$ samples, each of size $n-k$ from the normal distribution $N(\mu, \sigma^2)$ ($X_{1,i}, \dots, X_{n-k,i} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ for each $i = 1, \dots, 1000$), and for having k outliers in the sample data, we generate 1000 samples, each of size k from the normal distribution with different parameters ($X_{n-k+1,i}, X_{n,i} \sim N(\mu+\delta, \sigma^2)$ for each $i = 1, \dots, 1000$). Therefore, the MSE values of the proposed parametric and robust estimators of S_{pmk} and C''_{pmk} based on the sample data $\{X_{1,i}, \dots, X_{n-k,i}, X_{n-k+1,i}, \dots, X_{n,i}\}$ for $i = 1, \dots, 1000$ are calculated by the following formulas in this investigation:

- **Computation of the MSE values of robust estimators:** The MSE criterion for the IQR-based and MAD-based estimators of S_{pmk} and C''_{pmk} for the considered sample of size n in the presence of k outliers is computed using the following formulas:

$$\text{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{S}^{\text{IQR}}_{pmk,i}-S_{pmk}\right)^2,$$

$$\text{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{S}^{\text{MAD}}_{pmk,i}-S_{pmk}\right)^2,$$

$$\text{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{C}''^{\text{IQR}}_{pmk,i}-C''_{pmk}\right)^2,$$

and

$$\text{MSE} = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{C}''^{\text{MAD}}_{pmk,i}-C''_{pmk}\right)^2,$$

TABLE 1 | Results of the MSE of the PCI S_{pmk} based on $N(\mu = 3, \sigma^2 = 4^2)$, for various estimators with different numbers of outliers on the basis of the preset SLs ($LSL = 0, T = 6, USL = 10$).

		Robust procedure	Robust procedure	Parametric procedure	Parametric procedure
n	***k***	**IQR-based estimator**	**MAD-based estimator**	**MM-based estimator**	**ML-based estimator**
16	1	0.0137	0.0089	0.0049	0.0047
	2	0.0123	0.0111	0.0048	0.0048
	3	0.0130	0.0123	0.0048	0.0058
22	1	0.0129	0.0061	0.0033	0.0032
	2	0.0128	0.0066	0.0036	0.0037
	3	0.0116	0.0069	0.0034	0.0038
28	1	0.0127	0.0044	0.0024	0.0023
	2	0.0115	0.0056	0.0026	0.0025
	3	0.0111	0.0052	0.0024	0.0025
34	1	0.0127	0.0039	0.0020	0.0020
	2	0.0125	0.0038	0.0020	0.0020
	3	0.0120	0.0042	0.0021	0.0022
40	1	0.0126	0.0035	0.0017	0.0017
	2	0.0119	0.0033	0.0018	0.0018
	3	0.0119	0.0038	0.0018	0.0019
75	1	0.0127	0.0017	0.0009	0.0009
	2	0.0128	0.0017	0.0009	0.0009
	3	0.0126	0.0017	0.0009	0.0010
100	1	0.0127	0.0013	0.0007	0.0007
	2	0.0126	0.0013	0.0007	0.0007
	3	0.0122	0.0014	0.0007	0.0007
136	1	0.0133	0.0008	0.0005	0.0005
	2	0.0127	0.0009	0.0005	0.0005
	3	0.0125	0.0009	0.0005	0.0005

where S_{pmk} and C''_{pmk} are calculated using Equations (2) and (3), respectively, and also $\widehat{S}^{\mathrm{IQR}}_{pmk,i}$, $\widehat{S}^{\mathrm{MAD}}_{pmk,i}$, $\widehat{C}''^{\mathrm{IQR}}_{pmk,i}$, and $\widehat{C}''^{\mathrm{MAD}}_{pmk,i}$ for $i = 1, \dots, 1000$ are computed by Equations (6)–(9), respectively.

- **Computation of MSE values of the parametric estimators:** The MSE criterion for the MM- and ML-based estimators of S_{pmk} and C''_{pmk} for a given sample of size n with k outliers is determined using the following formulas:

$$\mathrm{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{S}^{MM}_{pmk,i} - S_{pmk} \right)^2,$$

$$\mathrm{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{S}^{ML}_{pmk,i} - S_{pmk} \right)^2,$$

$$\mathrm{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{C}''^{MM}_{pmk,i} - C''_{pmk} \right)^2,$$

and

$$\mathrm{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{C}''^{ML}_{pmk,i} - C''_{pmk} \right)^2,$$

where S_{pmk} and C''_{pmk} are computed using Equations (2) and (3), respectively. Additionally, $\widehat{S}^{MM}_{pmk,i}$, $\widehat{S}^{ML}_{pmk,i}$, $\widehat{C}''^{MM}_{pmk,i}$, and $\widehat{C}''^{ML}_{pmk,i}$ for $i = 1, \dots, 1000$ are determined using Equations (18)–(21), respectively.

To clarify the steps involved in calculating the MSE values provided in Tables 1 and 2, we use Algorithm 1, which generates sample data sets based on the following parameters: $m = 1000$, $LSL = 0$, $USL = 10$, $T = 6$, $\mu = 3$, $\sigma = 4$, and $\delta = 2$. This is done for various sample sizes ($n = 16(6)40$ and $75, 100, 136$) with different numbers of outliers ($k = 1, 2, 3$).

For example, in the sixth column of Table 1, for $n = 16$ with the presence of two outliers ($k = 2$), we generate $m = 1000$ samples, each of size $n - k = 16 - 2 = 14$ from the normal distribution $N(\mu = 3, \sigma^2 = 4^2)$ $(X_{1,i}, \dots, X_{14,i} \overset{i.i.d.}{\sim} N(3, 4^2))$ for each

TABLE 2 | Results of the MSE of the index C''_{pmk} based on $N(\mu = 3, \sigma^2 = 4^2)$, for various estimators with different numbers of outliers on the basis of the preset SLs ($LSL = 0, T = 6, USL = 10$).

n	***k***	**Robust procedure IQR-based estimator**	**Robust procedure MAD-based estimator**	**Parametric procedure MM-based estimator**	**Parametric procedure ML-based estimator**
16	1	0.0499	0.0083	0.0053	0.0049
	2	0.0705	0.0122	0.0068	0.0056
	3	0.0699	0.0111	0.0068	0.0042
22	1	0.0314	0.0051	0.0033	0.0030
	2	0.0475	0.0086	0.0047	0.0041
	3	0.0487	0.0071	0.0045	0.0032
28	1	0.0287	0.0046	0.0028	0.0026
	2	0.0323	0.0047	0.0031	0.0027
	3	0.0328	0.0049	0.0031	0.0025
34	1	0.0253	0.0039	0.0023	0.0022
	2	0.0264	0.0039	0.0024	0.0022
	3	0.0300	0.0041	0.0025	0.0021
40	1	0.0216	0.0032	0.0019	0.0018
	2	0.0228	0.0033	0.0022	0.0020
	3	0.0248	0.0034	0.0022	0.0018
75	1	0.0127	0.0015	0.0010	0.0009
	2	0.0147	0.0017	0.0011	0.0010
	3	0.0144	0.0016	0.0011	0.0010
100	1	0.0112	0.0011	0.0008	0.0008
	2	0.0120	0.0012	0.0008	0.0008
	3	0.0121	0.0012	0.0008	0.0008
136	1	0.0094	0.0008	0.0006	0.0005
	2	0.0098	0.0008	0.0006	0.0005
	3	0.0101	0.0008	0.0006	0.0006

$i = 1, \ldots, 1000$), and for having two outliers in the sample data, we generate 1000 samples, each of size two from the normal distribution with different parameters ($X_{15,i}, X_{16,i} \sim N\left(\mu + 2, \sigma^2\right)$ for each $i = 1, \ldots, 1000$). Therefore, by performing Algorithm 1 based on employing the ML-based estimator of S_{pmk} ($\widehat{S}^{ML}_{pmk}$) on the basis of the sample data $\left\{X_{1,i}, \ldots, X_{14,i}, X_{15,i}, X_{16,i}\right\}$ for $i = 1, \ldots, 1000$, the MSE of $\widehat{S}^{ML}_{pmk}$ is calculated by the following formula in this investigation:

$$\text{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left(\widehat{S}^{ML}_{pmk,i} - S_{pmk}\right)^2 = 0.0048,$$

where S_{pmk} is calculated by the following:

$$S_{pmk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{10-3}{\sqrt{4^2 + (3-6)^2}}\right) + \frac{1}{2}\Phi\left(\frac{3-0}{\sqrt{4^2 + (3-6)^2}}\right)\right\} = 0.3083.$$

The outcomes from Tables 1 and 2 (Figures 1 and 2) indicate a significant influence of sample size on the MSE for the parametric and robust estimation procedures. Upon comparing these methodologies, it becomes evident that the optimal results stem from utilizing ML- and MM-based estimators. This is attributed to the minimal impact of varying outlier quantities on the MSE for these estimators. Consequently, in the context of the MSE value derived from this simulation study, one can infer that the proposed ML- and MM-based estimators are more suitable for estimation tasks involving the presence of outliers.

Remark 3.2. In general, robust methods typically require more computational effort because they involve iterative procedures, numerical solutions, or more complex algorithms to handle outliers or deviations from assumptions. Parametric methods, on the other hand, tend to be faster and computationally simpler because they rely on specific distributional assumptions and often have closed-form solutions. However, in this case, our proposed parametric estimators ($\widehat{S}^{MM}_{pmk}$, $\widehat{C}''^{MM}_{pmk}$, $\widehat{S}^{ML}_{pmk}$, and $\widehat{C}''^{ML}_{pmk}$) require more computational effort than the robust estimators ($\widehat{S}^{\text{IQR}}_{pmk}$, $\widehat{C}''^{\text{IQR}}_{pmk}$, $\widehat{S}^{\text{MAD}}_{pmk}$, and $\widehat{C}''^{\text{MAD}}_{pmk}$) because solving Equations (12)–(16) necessitates the use of numerical methods.

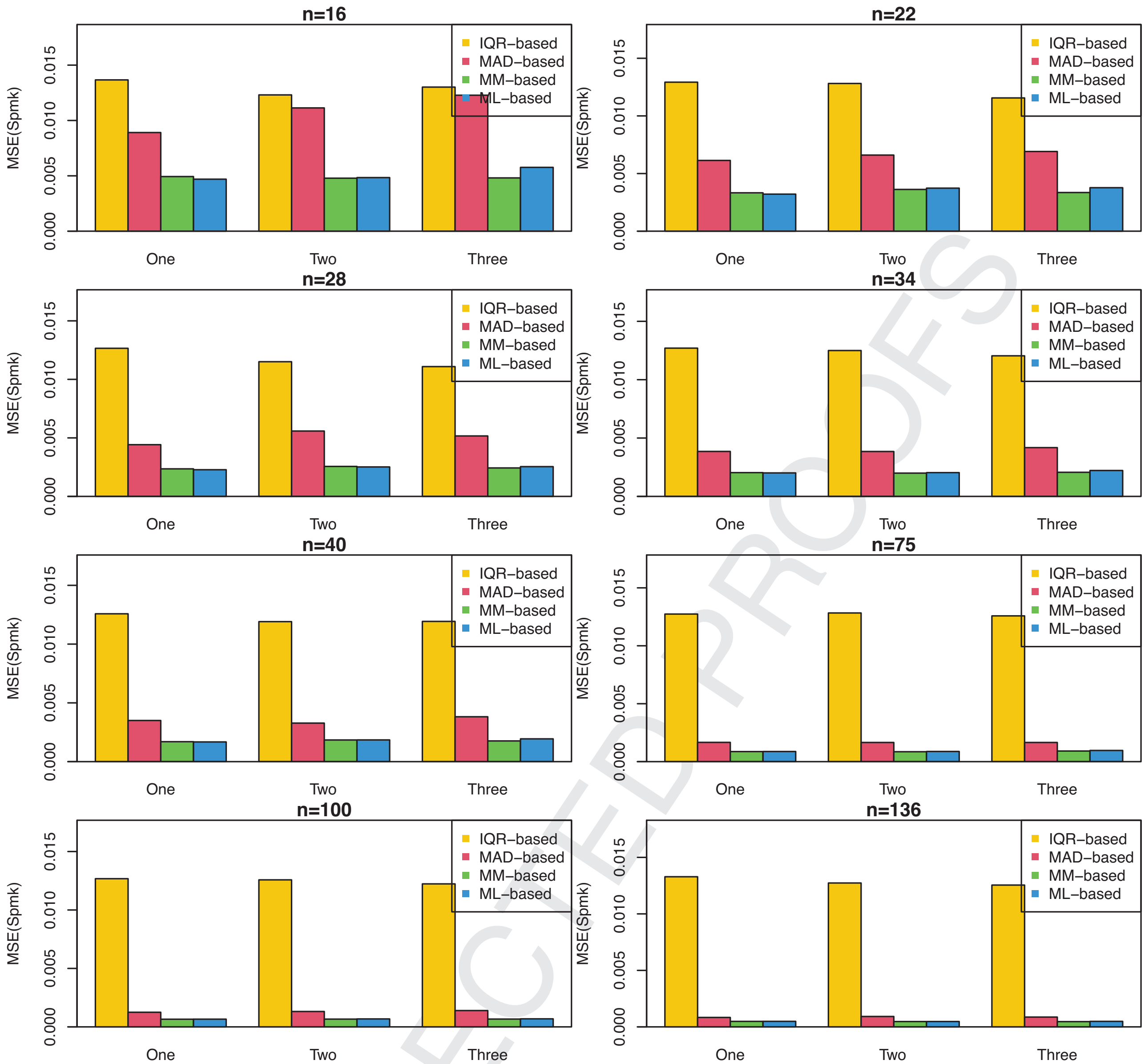

FIGURE 1 | The MSE of the index S_{pmk} for various estimators with different numbers of outliers ($k = 1, 2, 3$).

Remark 3.3. Note that the proposed procedure of the indices estimation is provided for the normal distribution of the process output. Herein, by inspiration of Slifker and Shapiro [25], we use a procedure based on sample percentiles for fitting Johnson distributions to data. This procedure is especially helpful when the collected data are non-normal, but one desires to apply a methodology that requires the underlying distribution to be normal. Under the non-normality assumption, to transform the variable X to a standard normal variable Z, we propose the Johnson's system of distributions which is generated by transformation of the form Slifker and Shapiro [25]

$$Z = \gamma + \eta f_i(X; \lambda, \epsilon), \tag{23}$$

where three functions $f_i(X; \lambda, \epsilon)$, for $i = 1, 2, 3$, associate with the Johnson's system. The parameters λ, η, and ϵ are estimated using the Johnson transformation procedure; see more details about Johnson's system of distributions in Bowman and Shenton [26], Hahn and Shapiro [27], Johnson et al. [28], Slifker and Shapiro [25], and Kendall et al. [29].

The following section presents the obtained results corresponding to the application of the novel estimations of PCIs with asymmetric tolerances to check the performance of an industrial manufacturing process using a real dataset.

4 | Illustrative Example and Interpretation of the Implications for Process Improvement

4.1 | Numerical Results

A real-world example is used in this section to demonstrate how well PCIs estimate when there are outliers. It is emphasized that, according to the definition, outliers are observations that deviate from the presumptive model.

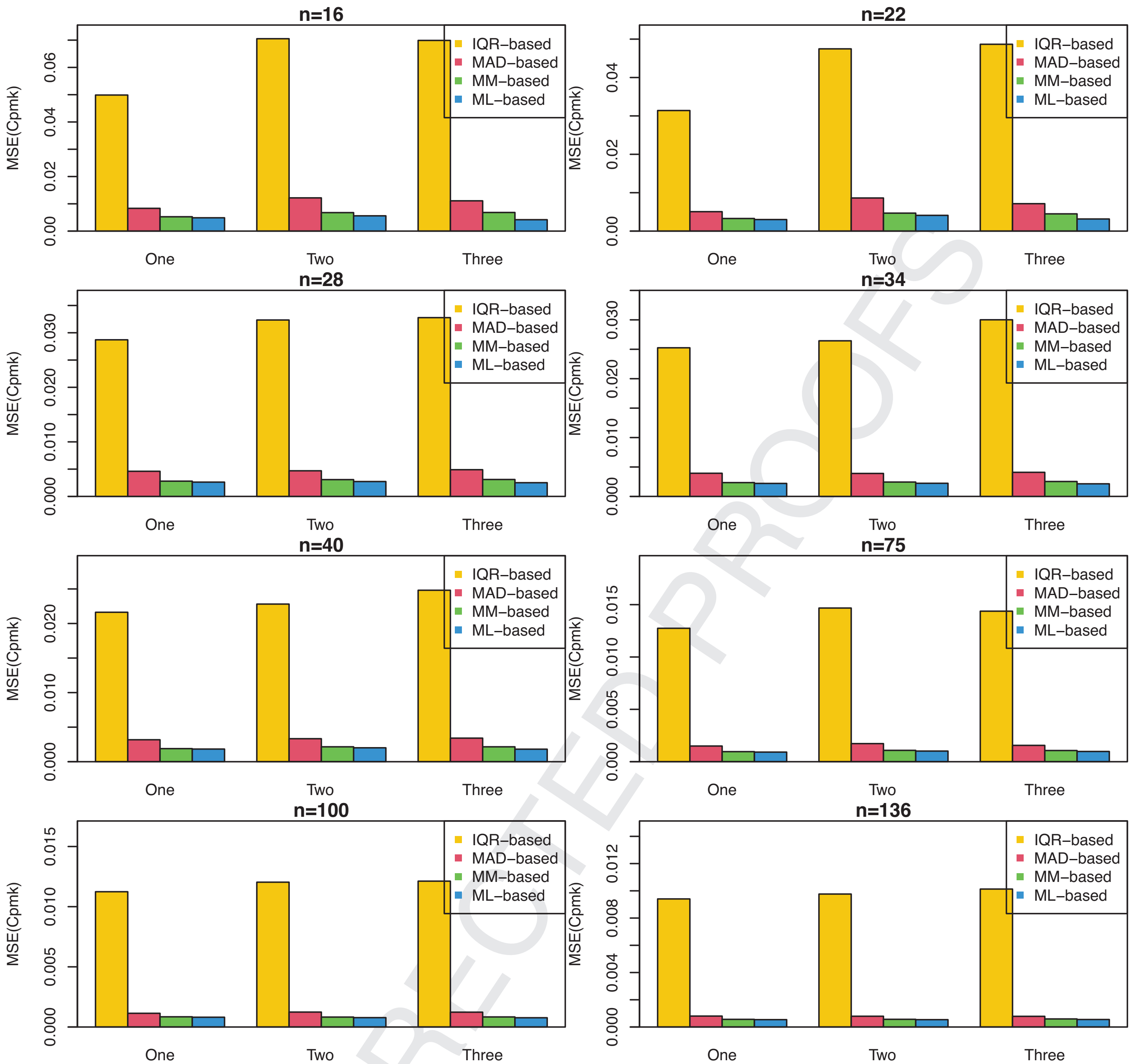

FIGURE 2 | The MSE of the index C''_{pmk} for different estimators with various numbers of outliers ($k = 1, 2, 3$).

An electronic circuit or device that amplifies the strength of a signal applied to its input is called an amplifier. Small signal amplifiers are frequently used devices in the field of electronics because of their capacity to amplify relatively small input signals, such as those from photo-devices or relays, into much larger output signals that can be used to power loudspeakers, lamps, or other devices. An ideal signal amplifier has the following three key characteristics: input resistance (R_{IN}), output resistance (R_{OUT}), and the gain of an amplifier. The gain of an amplifier refers to the increased difference between the input and output signals. Essentially, gain measures the extent to which an amplifier boosts the input signal. It can be expressed either in decibel (dB) or in numbers and represents how much an amplifier is able to amplify a signal given to it. Figure 3 indicates a type of an amplifier and Figure 4 shows an amplifier gain of the input signal.

The example presented in the following concerns with the capability of a process, which produces electronic telecommunication amplifiers. The following example relates to the capability of an electronic telecommunication amplifier production process. The Juran Institute [2] provides the original data. The gain amplifier is the quality characteristic of interest. The design of the amplifiers had called for a gain of 10 dB and allowed the amplifiers to be considered acceptable if the gain fell between 7.75 and 12.25 dB, that is, $(LSL, T, USL) = (7.75, 10, 12.25)$.

The quality improvement team measured the gains of 120 amplifiers as a sample to determine how capable the manufacturing process was that produced the amplifiers. Three common methods can be taken into consideration to examine the normality of data: (1) the Cullen and Frey graph, (2) the pdf plot, and (3) the

ALGORITHM 1 | Simulation procedure for computing MSE on the basis of generating m samples each of size n with k outliers.

Require:
(1) The desired index C_u and its considered estimator ($\widehat{C}_u$).
(2) $n \geq 1, k \geq 0, m \geq 1, \sigma > 0, \mu, \delta, LSL, USL \in \mathbb{R}$.
(3) Probability density for the random variable X.
(4) Probability density for the random outlier.
Ensure: MSE of the desired estimator $\widehat{C}_u$.

for $i = 1$ `to` m **do**
  `Generate independently` $X_{1,i}, \ldots, X_{n-k,i} \sim N(\mu, \sigma^2)$.
  `Generate independently` $X_{n-k+1,i}, \ldots, X_{n,i} \sim N(\mu + \delta, \sigma^2)$.
  `Combine two samples` $(X_{1,i}, \ldots, X_{n-k,i})$ `and` $(X_{n-k+1,i}, \ldots, X_{n,i})$,
  `to achieve` $\{X_{1,i}, \ldots, X_{n-k,i}, X_{n-k+1,i}, \ldots, X_{n,i}\}$.
  `Compute` $\widehat{C}_{u,i}$ `based on the sample data set` $\{X_{1,i}, \ldots, X_{n,i}\}$.
end for
`Calculate` C_u.
`Calculate MSE of the desired estimator` $\widehat{C}_u$ `by Equation (22)`.
return `MSE of the desired estimator` $\widehat{C}_u$.

FIGURE 3 | Schematic of a type of an amplifier.

goodness-of-fit test. Another name for the Cullen and Frey graph is the skewness-kurtosis graph; see Cullen et al. [30]. It suggests choosing the best fit based on kurtosis and skewness level for an unknown distribution. To aid in the selection of the model, values for common distributions are displayed on this graph. The left side graphs in Figure 5 shows the distribution of observation has not a skewness of zero, but the kurtosis of observation model is close to the normal distribution. Therefore, by examining the Cullen and Frey graph of the original amplifier gain data in the left side graph of Figure 5, the normal distribution model is not suitable to fit the original data.

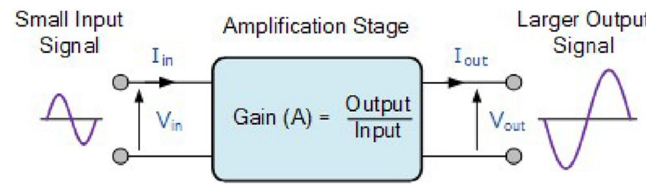

FIGURE 4 | Amplifier gain of the input signal.

TABLE 3 | The original amplifier gain data.

8.1	8.2	9.1	11.5	9.3	8.4	7.9	9.9	8.7	8.1	8.5	8.6
10.4	8.9	8.4	8.0	9.7	9.1	8.5	10.6	9.8	10.1	8.2	9.2
8.8	10.1	9.6	7.9	8.7	10.1	9.2	8.6	8.5	9.6	9.0	8.5
9.7	9.4	11.1	8.3	8.2	7.8	8.7	9.4	8.9	8.3	10.2	9.6
7.8	9.2	7.9	8.7	8.9	8.1	10.2	8.8	9.1	8.0	9.5	9.0
9.9	7.9	8.5	10.0	8.6	8.8	7.9	8.2	8.4	9.8	8.3	10.7
11.7	9.5	8.7	9.4	9.5	8.0	9.8	10.5	8.1	9.0	8.9	8.6
8.0	10.9	7.8	9.0	9.4	9.2	8.3	9.7	9.5	8.9	9.1	10.0
9.3	7.8	10.5	9.2	8.8	8.4	9.0	9.1	8.7	8.1	10.3	8.8
9.0	8.3	8.5	10.7	8.3	7.8	9.6	8.0	9.3	9.7	8.4	8.6

Shapiro–Wilk test does not confirm fitting the normal distribution on the original data with $W = 0.9523$ and p value $= 0.0003$. Also, the Lilliefors (corrected Kolmogorov–Smirnov) normality test with $D = 0.0815$ and p value $= 0.0487$, shows that the original amplifier gain data follow a non-normal distribution.

According to Remark 3.3, to transform the non-normal data to normality, we can use the Johnson transformation procedure. It must be noted that using the original specification limits, $(LSL, T, USL) = (7.75, 10, 12.25)$, to assess the quality using the transformed data would be incorrect. Therefore, the transformed specification limits $(LSL', T', USL') = (-2.314, 1.019, 6.302)$ as well as the transformed data are calculated by the following estimated transformation; see more details in [31]:

$$z = 0.96 + 0.98 \ln\left(\frac{x - 7.59}{4.68 + 7.59 - x}\right), \tag{24}$$

where z is the standard normal observation x.

Table 3 displays the sample of the original gains of 120 amplifiers listed in Juran Institute [2]. Table 4 displays the corresponding transformed amplifier gain data, using the estimated transformation in Equation (24).

Meanwhile, the transformed amplifier gain data set has been checked for the normal distribution's appropriateness using the goodness-of-fit test (Shapiro–Wilk test or Lilliefors normality test) and showed that the normal distribution is a suitable distribution to fit the transformed data. Also, based on the right-side graphs in Figure 5, the normal distribution model is appropriate to fit the transformed amplifier gain data.

The plots in Figure 6 display the box plots and histograms of the original and transformed amplifier gain data with the fitted density functions. Moreover, box plots and histograms show

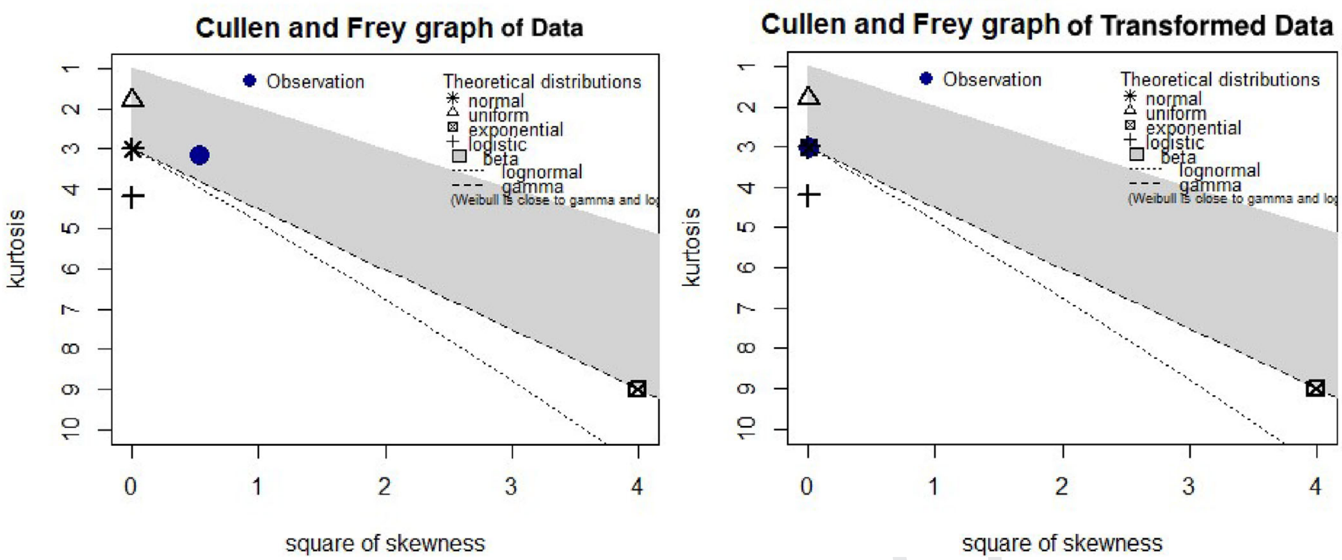

FIGURE 5 | Left: Cullen and Frey graph of the original amplifier gain data. Right: Cullen and Frey graph of the transformed amplifier gain data.

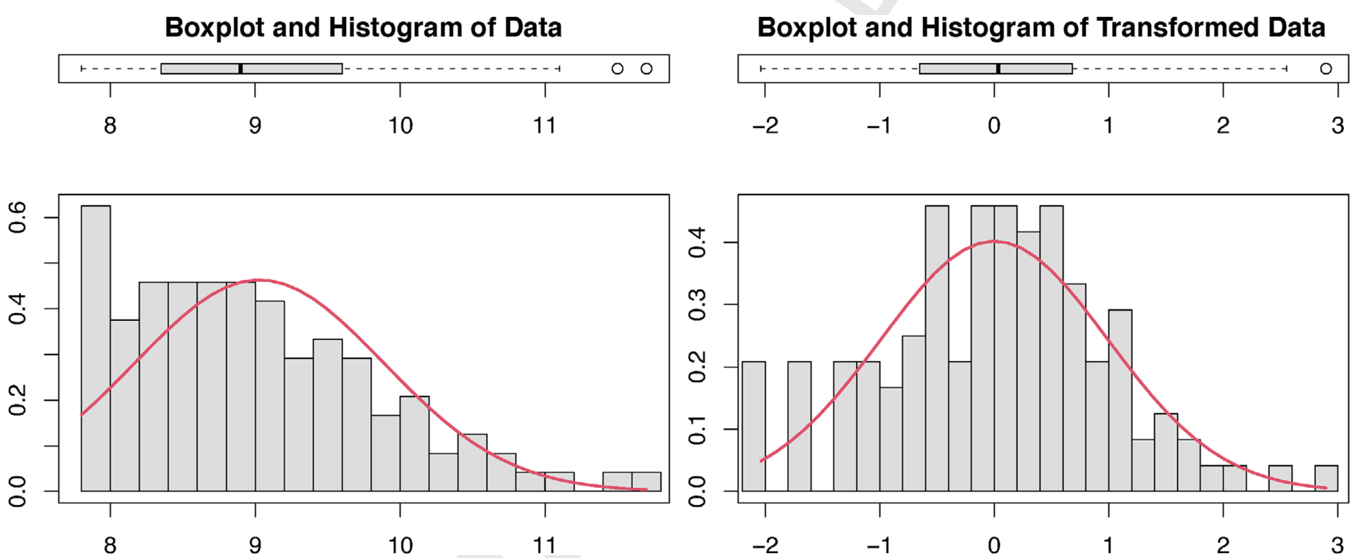

FIGURE 6 | Left: Histogram for 120 observed gain data. Right: Histogram for 120 transformed gain data.

that outliers contaminate both original and transformed data (Figure 6). Whenever the alternative hypothesis is that "highest value 2.9 is an outlier," the result of the chi-square test for the detection of the outlier in the transformed data shows that one outlier exists (p value = 0.003).

In the original dataset, extreme values might dominate and distort the overall distribution, making it harder to identify smaller but still significant outliers. After transformation, extreme values are often compressed (especially with log transformations), allowing outliers to stand out more clearly relative to the transformed distribution. It should be noted that the chi-square test is often more effective at detecting outliers in transformed data compared to the original dataset due to how the transformation changes the data distribution, enabling better adherence to the test's assumptions. Transforming data makes it more suitable for statistical tests like the chi-square test by reducing skewness, stabilizing variance, and aligning the data more closely with the normal distribution, thereby improving outlier detection.

Remark 4.1. The Johnson transformation reshapes both non-normal data and asymmetric tolerances into a form that adheres to normal distribution assumptions [6]. By doing this, it ensures that the indices like S_{pmk} and C''_{pmk} remain valid, accurately reflecting process capability. This transformation effectively handles the skewness and asymmetry in the original dataset by stabilizing variance and ensuring that distances between the mean and specification limits are properly represented in the transformed space. After transformation, the specification limits and data are normalized, meaning that the distance between the transformed mean and the transformed limits is correctly represented. This allows for accurate calculations of this indices, even when the tolerances were asymmetric. It should be noted that this manufacturing process operates based on asymmetric

TABLE 4 | The transformed amplifier gain data.

−1.1	−0.9	0.2	2.6	0.4	−0.6	−1.6	0.9	−0.2	−1.1	−0.4	−0.3
1.4	0.0	−0.6	−1.3	0.8	0.2	−0.4	1.5	0.9	1.1	−0.9	0.3
−0.1	1.1	0.7	−1.6	−0.2	1.1	0.3	−0.3	−0.4	0.7	0.1	−0.4
0.8	0.5	2.0	−0.7	−0.9	−2.0	−0.2	0.5	0.0	−0.7	1.2	0.7
−2.0	0.3	−1.6	−0.2	0.0	−1.1	1.2	−0.1	0.2	−1.3	0.6	0.1
0.9	−1.6	−0.4	1.0	−0.3	−0.1	−1.6	−0.9	−0.6	0.9	−0.7	1.6
2.9	0.6	−0.2	0.5	0.6	−1.3	0.9	1.4	−1.1	0.1	0.0	−0.3
−1.3	1.8	−2.0	0.1	0.5	0.3	−0.7	0.8	0.6	0.0	0.2	1.0
0.4	−2.0	1.4	0.3	−0.1	−0.6	0.1	0.2	−0.2	−1.1	1.3	−0.1
0.1	−0.7	−0.4	1.6	−0.7	−2.0	0.7	−1.3	0.4	0.8	−0.6	−0.3

tolerance $(LSL', T', USL') = (-2.314, 1.019, 6.302)$. Thus, using the indices S_{pmk} and C''_{pmk} to measure the performance of this manufacturing process can be useful.

In the sense of parametric model of outliers, the likelihood functions with respect to k by using MM and ML estimators of the parameters are summarized in Table 5. In the case of both MM- and ML-based estimators of the PCIs, the likelihood functions are maximized at $k = 1$ (see Table 5). Therefore, according to Notation 2.2, taking $k = 1$ into account, the ML- and MM-based estimators for the indices S_{pmk} and C''_{pmk} under the conditions of $LSL' = -2.314$, $USL' = 6.302$, and $T' = 1.019$ are computed as follows: $\widehat{S}^{MM}_{pmk} = 0.5838$, $\widehat{S}^{ML}_{pmk} = 0.8122$, $\widehat{C}''^{MM}_{pmk} = 0.3997$, and $\widehat{C}''^{ML}_{pmk} = 0.5263$. Furthermore, the robust estimators for the indices S_{pmk} and C''_{pmk} are, respectively, obtained as $\widehat{S}^{\mathrm{IQR}}_{pmk} = 0.4749$, $\widehat{S}^{\mathrm{MAD}}_{pmk} = 0.6717$, $\widehat{C}''^{\mathrm{IQR}}_{pmk} = 0.5825$, and $\widehat{C}''^{\mathrm{MAD}}_{pmk} = 0.4909$.

The low values of the computed ML-based, MM-based, and robust estimators of the indices S_{pmk} and C''_{pmk} show that the average quality of the amplifiers deviates significantly from the target value, even though all 120 amplifiers met the specification limits. Now that the manufacturing line was unable to produce amplifiers with average quality that were closer to the target value, the quality improvement team's investigation could concentrate on that reason.

4.2 | Interpretation of the Implications for Process Improvement

The given results demonstrate that the indices S_{pmk} and C''_{pmk}—calculated using the ML-based, MM-based, and robust estimators—show low values, indicating that the average quality of the amplifiers deviates significantly from the target value, even though all 120 amplifiers met the specification limits. This suggests a process issue where average performance does not align with the target, and the presence of outliers likely plays a crucial role in these deviations.

In the presence of outliers, the implications for process improvement become more complex and nuanced, as traditional methods of quality assessment may be distorted by such anomalies. The proposed robust estimators, designed to handle outliers, suggest that outliers may be having a substantial impact on the process, requiring more focused investigation. Meanwhile, the proposed ML- and MM-based estimators are designed on the basis of parametric methods. Parametric estimators are typically more efficient when the underlying assumptions (like normality and independence) are valid. They make use of full distributional information, providing precise estimates when these assumptions hold true. However, these advantages can turn into limitations when the underlying assumptions do not hold, which is where robust estimators are preferred. Robust estimators perform better in the presence of outliers, skewed distributions, or non-normality. Here is a detailed interpretation of the implications for process improvement in this context:

- **Identification of variability:** The low values of these proposed estimators suggest that the process, while controlled within specifications, exhibits variability that prevents the amplifiers from consistently meeting the target quality. This signals a potential issue with process centering or alignment, even though amplifiers are within the acceptable performance range.
- **Focus beyond specifications:** Meeting specifications is necessary but not sufficient for process excellence. The fact that all amplifiers met the specification limits yet deviated from the target implies that quality improvements need to go beyond mere compliance with specifications. The focus should shift towards reducing variation and increasing precision toward the target value.
- **Root cause analysis:** The quality improvement team can now focus on why the average quality is not aligning with the target.
- **Opportunities for refinement:** To bring the process closer to the target, refinements might be needed:
  1. **Process optimization:** SPC tools, such as process capability studies, could help pinpoint where and why the process drifts from the target.
  2. **Continuous improvement (CI) initiatives:** Techniques like Six Sigma or Kaizen could be employed to minimize variability and ensure that production more consistently hits the target quality.
  3. **Employee training:** If the issue stems from manual interventions or human error, staff training or procedural changes might be necessary.
- **Monitoring and feedback:** Ongoing monitoring using robust quality control methods is crucial. The low values in the indices suggest that historical monitoring might not have been sensitive enough to capture these deviations, so improved data collection and real-time monitoring might be needed.
- **Customer satisfaction:** Deviations from the target, even if within specifications, might affect downstream performance or customer satisfaction. If the target quality reflects optimal product performance, failing to consistently achieve this could lead to reduced reliability, performance issues, or increased returns over time.
- **Robustness and scalability:** The use of MM-based, ML-based, and robust estimators indicates an advanced analysis of the data, suggesting that the current quality framework

TABLE 5 | The likelihood function with respect to the ML- and MM-based estimators of the parameters for different values of k in the illustrative example.

Procedure	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
MM	$4.114462e^{-76}$	$6.378718e^{-83}$	$1.566693e^{-81}$	$7.567683e^{-93}$	$2.1674493e^{-123}$
ML	$1.801374e^{+57}$	$0.921673e^{+44}$	$0.831332e^{+30}$	$0.450021e^{+22}$	$0.1304702e^{+12}$

might be effective but underutilized. This analysis can provide a roadmap for improving the robustness of the production process, making it more scalable and resilient to future demands.

5 | Conclusions

PCIs serve as effective tools for assessing the capability of a process in a controlled state. There has been extensive discussion about how the existence of outliers can negatively impact statistical analyses and decision-making processes. In this regard, we first reviewed some existing generalizations of the index C_{pmk} with asymmetric tolerances to check the performance of an industrial manufacturing process. Then, for processes with normal distributions, we proposed new parametric and robust estimators of the indices S_{pmk} and C''_{pmk} to estimate and compare these indices in the presence of different outliers. The results presented in Tables 1 and 2 indicated that the parametric estimators for indices S_{pmk} and C''_{pmk} perform better than the robust estimators in this context. Furthermore, the results were visually depicted in Figures 1 and 2. The mean squared error of these estimators exhibits a decreasing trend in relation to the sample size. Real data analysis was used to further clarify the proposed procedure. As a potential procedure for the future research, one can use these proposed robust/parametric estimators for the newly proposed robust quality test. These guidelines can be used to evaluate the capability of processes that involve multiple characteristics.

Acknowledgments

The authors would like to thank the respected anonymous referees whose comments improved the paper.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. S. Kotz and N. L. Johnson, "Process Capability Indices—A Review, 1992-2000," *Journal of Quality Technology* 34, no. 1 (2002): 2–19.

2. Juran Institute. "The Tools of Quality, Part IV: Histograms," *Quality Progress* 9 (1990): 75–78.

3. V. E. Kane, "Process Capability Indices," *Journal of Quality Technology* 18, no. 1 (1986): 41–52.

4. R. A. Boyles, "Process Capability With Asymmetric Tolerances," *Communications in Statistics-Simulation and Computation* 23, no. 3 (1994): 615–635.

5. K. Vännman, "A General Class of Capability Indices in the Case of Asymmetric Tolerances," *Communications in Statistics-Theory and Methods* 26 (1997): no. 8, 2049–2072.

6. W. L. Pearn, P. C. Lin, and K. S. Chen, "On the Generalizations of the Capability Index C_{pmk} for Asymmetric Tolerances," *Far East Journal of Theoretical Statistics* 3, no. 1 (1999): 47–66.

7. W. L. Pearn, P. C. Lin, and K. S. Chen, "Estimating Process Capability Index C_{pmk} for Asymmetric Tolerances: Distributional Properties," *Metrika* 54, no. 3 (2002): 261–279.

8. H. Iranmanesh, A. Parchami, and B. Sadeghpour Gildeh, "A Case Study on Quality Test Based on Fuzzy Specification Limits," in *International Conference on Intelligent and Fuzzy Systems* (Cham, Switzerland: Springer, 2021), 636–643.

9. H. Iranmanesh, A. Parchami, and B. Sadeghpour Gildeh, "Statistical Testing Quality and Its Monte Carlo Simulation Based on Fuzzy Specification Limits," *Iranian Journal of Fuzzy Systems* 19, no. 3 (2022): 1–17.

10. H. Iranmanesh, A. Parchami, and M. Jabbari Nooghabi, "Testing Capability Index C_{pk} With Its Application in Automobile Engine Manufacturing Industry," *Quality Engineering* 35, no. 1 (2023): 48–55.

11. A. Parchami, H. Iranmanesh, and B. Sadeghpour Gildeh, "Simulation Testing of Fuzzy Quality with a Case Study in Pipe Manufacturing Industries," in *International Conference on Intelligent and Fuzzy Systems* (Cham, Switzerland: Springer, 2021), 630–635.

12. A. Parchami, H. Iranmanesh, B. Sadeghpour Gildeh, "Monte Carlo Statistical Test for Fuzzy Quality," *Iranian Journal of Fuzzy Systems* 19, no. 1 (2022): 115–124.

13. K. P. Lu and S. T. Chang, "Robust Algorithms for Multiphase Regression Models," *Applied Mathematical Modelling* 77 (2020): 1643–1661.

14. S. Mehri, M. M. Ahmadi, H. Shahriari, and A. Aghaie, "Robust Process Capability Indices for Multiple Linear Profiles," *Quality and Reliability Engineering International* 37, no. 8 (2021): 3568–3579.

15. H. Iranmanesh, M. Jabbari Nooghabi, and A. Parchami, "Robust Yield Test for a Normal Production Process," *Quality Engineering* 36, no. 2 (2024): 273–286.

16. S. Prasad and T. Bramorski, "Robust Process Capability Indices," *Omega* 26, no. 3 (1998): 425–435.

17. M. Jabbari Nooghabi, "Process Capability Indices in Normal Distribution With the Presence of Outliers," *Journal of Applied Statistics* 47, no. 13-15 (2020): 2443–2478.

18. U. J. Dixit, "Characterization of the Gamma Distribution in the Presence of k Outliers," *Bulletin Bombay Mathematical Colloquium* 4 (1987): 54–59.

19. U. J. Dixit and M. Jabbari Nooghabi, "Efficient Estimation in the Pareto Distribution With the Presence of Outliers," *Statistical Methodology* 8, no. 4 (2011): 340–355.

20. U. J. Dixit and M. Jabbari Nooghabi, "Efficient Estimation of the Parameters of the Pareto Distribution in the Presence of Outliers," *Communications for Statistical Applications and Methods* 18, no. 6 (2011): 817–835.

21. F. R. Hampel, "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association* 69 (1974): 383–393.

22. P. J. Rousseeuw and C. Croux, “Alternatives to the Median Absolute Deviation,” *Journal of the American Statistical Association* 88, no. 424 (1993): 1273–1283.

23. J. Yang, S. Rahardja, and P. Fränti, “Outlier Detection: How to Threshold Outlier Scores?,” in *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing* (Sanya, China, 2019), 1–6.

24. K. Soetaert and M. K. Soetaert, “Package “rootSolve”,” (2023).

25. J. F. Slifke and S. S. Shapiro, “The Johnson System: Selection and Parameter Estimation,” *Technometrics* 22 (1980): 239–246.

26. K. O. Bowman and L. R. Shenton, “Johnson’s System of Distributions,” *Encyclopedia of Statistical Sciences* 4 (1983): 303–314.

27. G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering* (Hoboken: Wiley, 1994).

28. N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1 (New York, NY: John Wiley & Sons, 1994).

29. M. G. Kendall, A. Stuart, J. K. Ord, eds., *Kendall’s Advanced Theory of Statistics* (New York: Oxford University Press, 1987).

30. A. C. Cullen, H. C. Frey, and C. H. Frey, *Probabilistic Techniques in Exposure Assessment*, 1st ed. (New York: Springer Science & Business Media, 1999).

31. Y. M. Chou, A. M. Polansky, and R. L. Mason, “Transforming Non-Normal Data to Normality in Statistical Process Control,” *Journal of Quality Technology* 30, no. 2 (1998): 133–141.