Jurnal Teknologi Informasi dan Pendidikan
Volume 18, No. 1, March 2025
https://doi.org/10.24036/jtip.v18i1.962

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Potential of Robust Regression Methods in Clock Skew Measurement

Putu Ayu Citra Setiawan[1*]✉, Komang Oka Saputra[1], Dewa Made Wiharta[1]
[1]Master of Electrical Engineering, Faculty of Engineering, Udayana University, Badung, Indonesia
Corresponding Author: ptayucitra@gmail.com

Article Information

Article history:

No. 962
Rec. May 23, 2025
Rev. June 24, 2025
Acc. June 25, 2025
Pub. June 27, 2025
Page. 794 – 810

Keywords:

- *Clock Skew*
- *RANSAC*
- *Least Median of Squares*
- *S-Estimator*
- *Robust Regression*

ABSTRACT

Clock skew, defined as the difference in clock rates between digital devices, serves as a unique and stable fingerprint for device identification and authentication, particularly in distributed network environments. Traditional clock skew estimation techniques, such as linear regression, are effective under stable conditions but often fail in the presence of data disturbances, such as latency, jitter, and asymmetric delays, which introduce outliers. This study explores the application of robust regression methods to enhance the accuracy and stability of clock skew estimation under such conditions. Three robust techniques are comparatively analyzed: Least Median of Squares (LMedS), Random Sample Consensus (RANSAC), and S-Estimators. LMedS offers high resistance to outliers by minimizing the median of squared residuals, though it is computationally demanding for large datasets. RANSAC achieves a practical balance between robustness and efficiency through iterative model fitting and inlier maximization, while S-Estimators provide strong statistical resistance to both outliers and high-leverage points, albeit with increased implementation complexity. The comparative evaluation considers key parameters such as estimation accuracy, computational cost, and robustness to anomalies. Results indicate that RANSAC is generally preferred for clock skew measurement in distributed systems due to its efficient performance and explicit outlier detection capabilities. However, LMedS and S-Estimators remain valuable in scenarios with more complex anomaly structures or higher noise levels. This study contributes to the selection of appropriate robust regression methods for reliable clock skew estimation in dynamic and error-prone network environments.

How to Cite:

Setiawan, P. A. C., & et al. (2025). Potential of Robust Regression Methods in Clock Skew Measurement. Jurnal Teknologi Informasi Dan Pendidikan, 18(1), 794-810. https://doi.org/10.24036/jtip.v18i1.962

1. INTRODUCTION

Clock skew is defined as the difference in clock rates between digital devices, reflecting deviations from real time. It serves as a stable fingerprint that helps distinguish one device from another [1]. Typically measured in parts per million (ppm) or even higher precision, clock skew exhibits device-specific characteristics, making it a valuable metric for device fingerprinting [2]. This metric is particularly important in the context of security and device identification, as the unique clock skew differences can be used to recognize and authenticate devices in distributed networks.

Clock skew is often visualized as the slope of an offset-set diagram, which depicts the relationship between the time recorded by a device and the time measured by an external reference, such as UTC time. In ideal conditions, where clock skew is stable and devices communicate through controlled channels, the data points on this diagram will be distributed along a straight line. This makes linear regression an effective method for estimating clock skew, provided the data is not contaminated by external disturbances.

However, in practice, clock skew measurements are not always taken under stable conditions. The process of collecting timestamps often involves delays or latency between the measuring device and the device being measured. Therefore, a stable environment, such as wired communication, is essential for ensuring the accuracy of the collected data. In contrast, in less stable conditions, such as wireless networks, these delays can cause shifts in the offsets, and the resulting data often leads to a deviation from the densest region in the diagram. Consequently, outliers or non-representative data points may emerge, obscuring the accurate estimation of clock skew.

Several methods have been developed to measure clock skew, aiming to enhance the accuracy of device identification and time synchronization. [3] One approach involves constructing parallelograms from coexisting angle-region tuples observed at the beginning and end of the offset set, thereby improving the capability of the Hough transform-based method to effectively detect jump points. Meanwhile, [4] introduces the Offset Reconstruction Method to restore outliers to their original positions, resulting in an ideal dataset for precise clock skew measurement. Additionally, by applying local weighting and progressive compensation, [5] addresses the underfitting problem commonly encountered in linear regression, while employing Least Trimmed Squares (LTS) to mitigate the influence of outliers on regression outcomes. This approach has proven effective in minimizing the impact of outliers on synchronization accuracy. LTS is a robust regression method designed for situations where the distribution of residuals is non-normal or when the data contains influential outliers that can affect the model [6]. Although the LTS estimator is well known for its high robustness against the presence of outliers, it is characterized by relatively low efficiency [7].

Building upon previous advancements, this paper explores several promising approaches for effective clock skew estimation, with a focus on robust regression methods to achieve stable and accurate measurements. This study analyzes three potential robust regression methods that have not been previously applied to clock skew estimation, with maximum breakdown point (BP) 50%. The Breakdown Point (BP) is a global measure of an estimator's robustness, representing the maximum proportion of gross errors or outliers in a dataset that the estimator can withstand before producing arbitrarily large or incorrect results [8]. In other words, it quantifies the estimator's ability to maintain reliability in the presence of data contamination. A higher breakdown point indicates that the estimator is more resilient to the influence of extreme values, making it particularly suitable for applications involving noisy or corrupted data. First method is Least Median of Squares (LMedS). LMedS is a parameter estimation method that seeks to minimize the median of the squared residuals, making it highly resistant to outliers with a breakdown point of up to 50% [9]. This approach is particularly effective when applied to data with high levels of noise or extreme disturbances, as commonly encountered in applications such as computer vision and signal processing [10], [11], [12].

The second method is Random Sample Consensus (RANSAC). Since its introduction, RANSAC has become a standard and widely used technique for detecting outliers and estimating model parameters in datasets with significant noise or contamination [13]. Its core strength lies in its ability to iteratively search for the best-fitting model by selecting random subsets of data and evaluating their consistency with the overall dataset. RANSAC performs particularly well in estimating the homography, especially in scenarios where scenes contain a high proportion of mismatches or outliers [9], [14], [13], [15], [16]. Homography is a method used by computers to transform or adjust images captured from different viewpoints so that they appear as if taken from the same position. The purpose of this transformation is to enable image alignment, such as in panoramic photo stitching, perspective correction, or planar object tracking in video sequences [17]. This capability makes RANSAC especially valuable in computer vision tasks such as image stitching, where precise alignment of overlapping images is essential for producing seamless and visually coherent results.

The last method is S-Estimator. Rousseeuw and Yohai [18] introduced the concept of S-estimators as a robust statistical method designed to estimate regression parameters, such as slope intercept, by minimizing a robust measure of the scale of residuals. Unlike the conventional least squares approach, which minimizes the variance of residuals and is highly sensitive to outliers, S-estimators aim to enhance robustness by employing alternative scale measures that are less influenced by extreme values. S-estimators provide a more reliable regression and framework when dealing with contaminated data or data that deviate from the assumptions of normality. This makes them particularly well-suited for situations where data integrity is compromised and classical methods fail to yield stable and accurate parameter estimates. In the study by [19], the S-Estimator demonstrated a significant advantage in handling data containing turbulence or high-leverage outliers. High-leverage outliers are data points that not only have extreme values in the independent

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

variables (high leverage) but also deviate significantly from the general trend of the data (outliers). These points can exert disproportionate influence on the regression model, potentially leading to biased parameter estimates, distorted model fit, and misleading conclusions [20]. Unlike the Ordinary Least Squares (OLS) and Huber Maximum Likelihood Estimator, which are susceptible to disturbances caused by leverage and tend to break down, the S-Estimator maintains stable and reliable performance [21].

To deepen the understanding of the effectiveness of robust regression methods in clock skew estimation, it is essential to consider various parameters such as efficiency, computational complexity, and accuracy in handling data with outliers. Each of these parameters plays a crucial role in selecting the most suitable method for a particular application, especially in the context of networks prone to data disturbances such as packet loss, jitter, or asymmetric delays. By comprehensively analyzing and comparing these parameters, this study aims to identify the most suitable robust regression method for implementing clock skew estimation in complex network environments. Given the dynamic and unpredictable nature of modern networks, where factors such as packet loss, jitter, and asymmetric delays can significantly affect data integrity, selecting a method that balances robustness, efficiency, and computational feasibility is crucial. By understanding the strengths and trade-offs of each method, the findings of this study are expected to provide valuable insights for optimizing clock skew estimation models.

2. RESEARCH METHOD

This study employs a descriptive qualitative research approach to examine several emerging robust regression-based methods that show potential for implementation in clock skew estimation. The methodology is inspired by the frameworks presented in [1], [2], [22], and consists of several key stages: literature collection, data selection and categorization, theoretical analysis, synthesis and discussion, and conclusion formulation. This study does not utilize empirical data, but is instead based entirely on secondary literature obtained from various scholarly sources, including journals, conference proceedings, and academic articles. The primary objective is to review and compare robust regression methods developed in previous research.

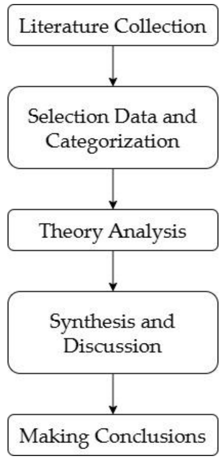

Figure 1. Research Methodology

In Literature Collection stage, A thorough search is conducted in well-established academic databases such as IEEE Xplore, SpringerLink, Google Scholar, and ScienceDirect. Keywords such as "clock skew estimation,", "robust regression", "Least Median Squares”, “LMeds", "RANSAC," and "S-Estimator" are used to identify literature that addresses methods with potential applicability for clock skew measurement. Articles are then screened for relevance based on their abstracts, titles, and keywords, ensuring they align with the study's focus. The collected scholarly articles are those published within the last 5 years (2020-2025). Once the literature has been collected, data selection is conducted according to the following criteria.

Table 1. Inclusion and Exclusion Criteria

Stage	Inclusion Criteria	Exclusion Criteria
Initial Stage	- According to the search keyword - Published 2020 -2025 - Proceeding or journal	- Published outside 2020 –2025
Stage-1 (Selection of title and abstract)	Methods that are relevant to robust regression, LMeds, RANSAC and S-Estimator.	Methods that are not relevant to robust regression, LMeds, RANSAC and S-Estimator
Stage-2 (Full-Text Assessment)	- Utilizing LMeds, RANSAC, or S-Estimator as the primary approach. - Must include quantitative metrics (e.g., RMSE, breakdown point, bias) and comparisons with other methods. - A clear description of the dataset (simulated/real) and the method parameters.	The paper contains only theoretical/ mathematical analysis without experimentation.

Following the selection process, the chosen articles were systematically classified into several thematic categories aligned with the primary focus of the research. These

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

categories include a comprehensive discussion of the methodologies to be compared namely, including Least Median of Squares (LMedS), RANSAC, and S-Estimator along with an analysis of their respective potentials for application in clock skew estimation.

In the theoretical analysis phase, a comprehensive evaluation was conducted to assess the potential application of the including Least Median of Squares (LMedS), RANSAC and S-Estimator methods in clock skew estimation. This evaluation was based on a critical synthesis of relevant literature, with a focus on an in-depth comparison of the theoretical foundations of each method, their robustness to outliers, computational efficiency, and overall effectiveness.

In the synthesis and discussion phase, the results obtained from the analysis of various scholarly sources were systematically consolidated to identify interrelationships among findings and to evaluate the consistency, applicability, and relevance of the investigated methods in the context of clock skew estimation. This synthesis aimed to develop a cohesive understanding of the relative strengths and limitations of each methodological approach. Specifically, methods such as Least Median of Squares (LMedS), RANSAC, and the S-Estimator were critically examined in the discussion section with regard to contextual factors such as robustness to outliers and computational efficiency. The synthesis process was conducted scientifically by first categorizing the selected studies according to methodological similarities, application domains, and evaluation criteria. Each method was then analyzed in terms of its robustness, underlying assumptions, and suitability for clock skew estimation. A comparative discussion followed to identify recurring patterns, methodological strengths, and existing research gaps, which form the basis for the conclusions and recommendations proposed in this study.

In the conclusion phase, the key findings gathered throughout the research process are integrated and consolidated to provide a comprehensive overview of the effectiveness, potential, and relevance of the methods analyzed in the context of clock skew estimation. Each finding is critically examined to identify the strengths and weaknesses of the methods employed, and to assess how these methods can be adapted to estimate clock skew.

3. RESULTS AND DISCUSSION

3.1. Least Median of Squares (LMedS)

The basic principle of the LMedS method is to fit a model that minimizes the median of the squared residuals, rather than the sum of the squared residuals as in Ordinary Least Squares (OLS) [23]. By doing so, LMedS places less weight on outliers, treating them as data points that do not conform to the underlying distribution of the majority of the data. This makes the LMedS method particularly useful in scenarios where data may contain extreme or anomalous values that would otherwise skew the parameter estimates.

$$MJ = \left(med\ \in_i^2\right) = \{M_1, M_2, \ldots\ldots, M_i\} \quad (1)$$

$\in_i$ denotes the residual, or the error between the observed value and the predicted value produced by the model. This residual reflects the deviation of the model from the actual data at the i-th observation. The term MJ refers to the minimum value of the median of squared residuals, obtained by evaluating multiple subsets of the data, and serves as the main criterion for selecting the most robust model parameters [24]. By minimizing the median rather than the mean, this method ensures that up to 50% of the data can be contaminated without significantly affecting the resulting model [7], [23], [25], [9]. This property makes LMedS particularly suitable for datasets where the presence of noise or gross errors cannot be ignored. Although it may sacrifice some statistical efficiency under ideal conditions, the trade-off is a substantial gain in robustness and resistance to outliers. The following are the steps of the LMedS (Least Median of Squares) estimation method for clock skew estimation [26] :

1) From the entire timestamp dataset, several random subsets are generated.
2) For each subset, regression is performed using the Ordinary Least Squares (OLS) method.
3) Residuals are calculated as the difference between the actual values and the predicted values from the model : $e_i = y_i - \hat{y}_i$.
4) For each model (obtained from a subset), the median of the squared residuals is computed : $median(e_i^2)$.
5) Among all subsets, the model with the lowest median of squared residuals is selected.
6) Data points with residuals below a certain threshold (inliers) are then reused to re-estimate the regression model.

In [10], the Least Median of Squares (LMedS) method is employed as a robust approach to enhance the accuracy and automation of radiometric intercalibration in DMSP-OLS night-time light imagery. The study introduces the LMedS-based Power Regression (LBPR) model, which demonstrates superior performance in terms of both accuracy—evidenced by the lowest adjusted RMSE—and computational efficiency when compared to other LMedS-based regression models, including linear, quadratic, exponential, and logarithmic forms. The application of LMedS enables stable calibration despite the presence of outliers, making it a reliable solution for large-scale remote sensing image preprocessing.

Similarly, in [11], the LMedS regression method was successfully applied to map land cover changes in Jakarta and its surrounding areas. Its robustness against outliers improved the accuracy of impervious surface classification, revealing that Jakarta and Tangerang exhibit higher impervious land coverage, whereas Bogor, Bekasi, and Depok remain predominantly non-impervious. This demonstrates the effectiveness of LMedS in handling urban land change analysis using remote sensing data.

In [12], robust regression utilizing the LMedS estimator was employed to analyze factors influencing soybean production in Indonesia, where the dataset contained outliers

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

violating normality assumptions. Due to its high breakdown point, LMedS effectively managed these outliers, resulting in more reliable parameter estimates. The study identified soybean seed quantity, field area, and rainfall as the most influential factors, with seed quantity exerting the strongest effect. This underscores the capability of LMedS to generate robust insights for agricultural policy-making despite non-ideal data conditions.

Collectively, these studies highlight the substantial potential of the Least Median of Squares method as a robust regression technique capable of delivering reliable results in the presence of outliers and noisy data across diverse applications. In the context of clock skew estimation, where measurement data are often contaminated with noise and outliers that can compromise synchronization accuracy, implementing LMedS could significantly enhance the robustness and reliability of skew estimation. By maintaining accurate estimates despite aberrant data points, LMedS presents a promising solution to minimize clock skew errors, thereby supporting the stability and security of time-critical communication and network systems.

3.2. Random Sample Consensus (RANSAC)

Random Sample Consensus (RANSAC) is an algorithm that employs random sampling to construct preliminary models from minimal subsets of data [16]. It iteratively refines these models by expanding them to better capture the underlying data distribution. RANSAC is particularly effective in simultaneously identifying inliers data points that fit the model well and outliers points that deviate significantly from the model. The iterative procedure in RANSAC is grounded in repeating the model estimation and evaluation process a predefined number of times, denoted as B iterations. In each iteration, a minimal random subset of the data is used to estimate a tentative model, which is then evaluated against the entire dataset to determine the number of inliers. Through this repetition, RANSAC increases the probability of identifying a model that accurately captures the underlying structure of the data, despite the presence of outliers. Notably, it has a high breakdown point (BDP > 50%), allowing it to tolerate a substantial proportion of outliers without degrading model accuracy [28].

The primary strength of this approach lies in its ability to converge toward a consensus model, a model that not only provides a robust fit to a significant portion of the data (the inliers), but also effectively identifies and excludes data points that deviate from the underlying distribution (the outliers). This capability makes RANSAC particularly well-suited for applications where datasets are heavily affected by noise or contain a substantial proportion of erroneous observations. RANSAC is capable of reducing the outlier mismatch rate to a manageable level. For instance, in offline scenarios, it can tolerate and still perform reliably with outlier ratios as high as 40% [9]. The standard RANSAC procedure for linear model fitting in clock skew estimation can be described through the following steps [14] :

1) Choose a small random subset S from the timestamp data. This subset contains a few data points $x(n_i)$, where each point is taken at a random position n_i.
2) Use the points in subset S to estimate the parameters $\hat{a}$ and $\hat{b}$ of a straight line model $x(n) = an + b$. This can be done by solving simple equations similar to linear regression using the selected points.
3) For all data points in the dataset, calculate how far each point is from the estimated line. The distance formula is :
$$d_n = \frac{|\hat{a}n + \hat{b} - x(n)|}{\sqrt{1 + \hat{a}n}} \quad (2)$$
4) If many points have a distance smaller than a chosen threshold t, they are marked as inliers and put into a new set D. Then, use all points in D to recalculate better values for a and b.
5) If the number of inliers is too small, repeat the steps with a new random subset. The process continues until enough inliers are found or a maximum number of iteration is reached.

Paper [9] introduces a linear relative pose estimation algorithm based on pose-only imaging geometry, which filters outliers through a reweighting strategy. To enhance robustness, the method integrates RANSAC with Iteratively Reweighted Least Squares (IRLS), enabling effective outlier rejection even under degenerate conditions—such as planar scenes—and with outlier ratios as high as 80%. The approach achieves up to a tenfold improvement in relative rotation accuracy, indicating strong potential for time-sensitive applications such as clock skew estimation in noisy environments.

In a related context, paper [14] proposes a sparsity-driven signal reconstruction framework that employs a customized RANSAC algorithm to identify reliable inliers from heavily corrupted signals. This method leverages the principle that disturbances degrade sparsity in the transform domain (e.g., the Discrete Fourier Transform) and requires no prior assumptions regarding the noise distribution. By incorporating RANSAC into the compressive sensing pipeline, the approach maintains high robustness without relying on specific statistical noise models.

A comparative analysis presented in paper [13] evaluates three outlier handling techniques: RANSAC, least squares, and Least Median of Squares (LMedS). The results consistently show that RANSAC provides superior robustness, particularly under high outlier conditions. This reinforces its relevance for applications that demand reliable estimation in the presence of noise and data contamination.

Paper [15] introduces GR-RANSAC, an extension of the classical RANSAC algorithm that incorporates geometric constraints based on 2D spatial relationships among feature points. By pre-filtering matches using distance and angular metrics, GR-RANSAC

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

improves the inlier ratio and reduces the number of required iterations, facilitating faster convergence. This suggests its viability for real-time or computationally constrained tasks.

Lastly, paper [16] presents PSSC-RANSAC, an accelerated RANSAC variant designed for aerial image registration under conditions of severe noise and complex transformations. It incorporates prior sampling strategies based on texture magnitude, spatial consistency, and feature similarity to prioritize high-quality samples and eliminate inconsistent subsets early in the process. The method demonstrates robustness against up to 90% outlier contamination, with notable reductions in computational overhead and reprojection error. These advancements further confirm RANSAC's adaptability and effectiveness in high-outlier environments.

Across all reviewed studies, RANSAC has demonstrated remarkable effectiveness in robust outlier detection, even in highly contaminated or sparse data environments. Its flexibility—through integration with reweighting schemes, geometric constraints, and prior-informed sampling—enables it to meet the accuracy and efficiency demands of modern signal processing tasks. These characteristics make RANSAC particularly well-suited for clock skew estimation, where reliable inlier selection amid noise is critical for accurate temporal analysis.

3.3. S-Estimator

The S-estimator is a robust estimation method characterized by a high breakdown point, which allows it to tolerate a substantial proportion of outliers without producing distorted results. However, it generally exhibits low statistical efficiency under ideal (i.e., normally distributed) conditions [30]. The S-estimator is typically derived by minimizing an M-estimator based on the residual scale. A notable limitation of M-estimators is their reliance on the median as a weighting measure, which reduces sensitivity to the overall data distribution and limits the use of available information. To address this issue, the S-estimator incorporates the standard deviation of the residuals or a robust alternative as a measure of scale, thereby enhancing its responsiveness to data dispersion and mitigating the weaknesses of M-estimation. Consequently, the S-estimator provides a more reliable approach for robust regression in the presence of outliers. The S-estimation method possesses a breakdown point of 50%, indicating that the presence of outliers in up to half of the observed data will not significantly affect the resulting regression model [7].

The key idea behind the S-estimator is the minimization of the dispersion of residuals, rather than the minimization of residual variance as in ordinary least squares [29]. Specifically, the S-estimator seeks the parameter vector β that minimizes a robust estimate of the scale of the residuals. Formally, the S-estimator is defined as the solution to the following minimization problem [31]:

$$\hat{\beta} = s(r_1(\hat{\beta}), r_2(\hat{\beta}), \dots, r_n(\hat{\beta})) \quad (3)$$

For a given sample, the S estimators can be obtained by solving the following equation, as proposed by Rousseeuw and Yohai [18]:

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_s}\right) = K \tag{4}$$

In this equation, K represents the expected value of the function ρ under the standard normal distribution. The term $\hat{\sigma}_s$ denotes the S-estimator of scale for the error term, and e_i are the residuals computed using the S-estimator of the regression coefficient vector β. In robust statistics, $\hat{\sigma}_s$ can be adapted using robust scale measures, such as the median absolute deviation (MAD), expressed as [30]:

$$\hat{\sigma}_i = \frac{median|e_i - median(e_i)|}{0.6745} \tag{5}$$

The median absolute deviation (MAD) measures the dispersion or variability of residual data by calculating the median of the absolute deviations of the residuals from their own median. The divisor 0.6745 is used to scale this estimator to be consistent under the assumption that the residuals follow a normal distribution. This value corresponds to a quantile of the standard normal distribution. The following presents an overview of the stages involved in the S-Estimator approach for clock skew estimation [30] :

1) Estimate the regression coefficients from the data using the Ordinary Least Squares (OLS) method.
2) Conduct classical assumption tests for the regression model to validate its applicability.
3) Detect the presence of outliers within the dataset.
4) Compute the initial coefficient estimates $\hat{\beta}^0$ using OLS.
5) Calculate the residuals $e_i = y_i - \hat{y}_i$
6) Calculate the scale estimator $\hat{\sigma}_i$ using formula (5).
7) Compute the standardized residuals $u_i = \frac{e_i}{\hat{\sigma}}$
8) Calculate the weights w_i as follows:
   - If $|u_i| \leq 1.547$ and it is the first iteration :

$$w_i = \left[1 - \left(\frac{u_i}{1.547}\right)^2\right]^2 \tag{6}$$

   - If $|u_i| > 1.547$and it is the first iteration :

$$w_i = 0$$

   - For later iterations, use the function :

$$w_i = \frac{\rho(u)}{u^2} \tag{7}$$

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

9) Re-estimate the regression coefficient $\hat{\beta}_s$ using the Weighted Least Squares (WLS) method with the computed weights w_i.
10) Repeat steps 5 to 8 until the coefficient estimates $\hat{\beta}_s$ converge.
11) Conduct hypothesis testing to assess whether the independent variable significantly affects the dependent variable.

Several recent studies have demonstrated the effectiveness of the S-Estimator in handling outliers and improving model robustness across various statistical contexts. In [21], the researchers examined the factors contributing to stunting across all districts and cities on Java Island using the Spatial Autoregressive (SAR) model. They found that the presence of outliers could significantly bias the estimation results. To address this issue, they implemented a combination of instrumental variables and the S-Estimator, which successfully improved the model's performance by lowering the residual standard error and increasing the R-squared value compared to models without the S-Estimator.

Similarly, in [32], the authors modeled the crime rate in Indonesia during the COVID-19 pandemic using macroeconomic indicators. Given that Ordinary Least Squares (OLS) regression is highly sensitive to outliers, they adopted robust regression with the S-Estimator, which produced a stable and reliable model. The model was able to explain 83.52% of the variance in crime rates, with significant predictors including unemployment, poverty, population density, and the human development index. The use of the S-Estimator in this context proved effective in mitigating the influence of extreme values that typically reduce estimation accuracy.

In [33], the focus shifted to semi-parametric estimators, such as matching methods, which were shown to be vulnerable to both bad and good leverage points. These outliers not only introduced bias but also disrupted the balance between covariates. To overcome this, the researchers proposed a reweighting strategy inspired by the Stahel-Donoho and multivariate S-Estimators, which effectively improved estimation performance in both simulation and real-world data.

Overall, these studies highlight the consistent advantages of using S-Estimators for robust parameter estimation in the presence of outliers. In the context of clock skew estimation, where the data are often influenced by noise, jitter, and temporal anomalies, the application of S-Estimator methods is particularly promising. Because of their ability to resist the distorting effects of outliers, S-Estimators can enhance the precision and stability of skew estimation algorithms. This is especially beneficial in systems where accurate timing is essential, such as in IoT networks and in-vehicle communication frameworks. Therefore, incorporating S-Estimator techniques into clock skew estimation frameworks may significantly improve their reliability and robustness under real-world, imperfect data conditions.

3.4. Discussion

In this section, we present a detailed review of the key architectural components and hyperparameter settings for each of the robust regression methods discussed previously. A thorough understanding of these critical hyperparameters and their tuning priorities is essential for the effective implementation and optimization of these methods in clock skew estimation. Table 2 summarizes the main hyperparameters associated with the Least Median of Squares (LMedS), Random Sample Consensus (RANSAC), and S-Estimator methods, along with the crucial aspects of their respective tuning processes. This comparison highlights how the performance and robustness of each method strongly depend on the careful selection and adjustment of these parameters, which directly influence convergence behavior, computational efficiency, and the capability to handle outliers within the data.

Table 2. Summary of Method Hyperparameters

Method	Key Hyperparameters	Tuning Focus
LMedS	Subset size, maximum iterations	Sufficient iterations, minimum subset to estimate model
RANSAC	Subset size, max iterations, residual threshold, min inliers	Residual threshold is critical
S-Estimator	Tuning constant, max iterations, initial scale estimate	Controls robustness and convergence behavior

Each robust regression method relies on specific hyperparameters that play a crucial role in balancing accuracy, robustness to outliers, and computational demands. For the LMedS method, the primary hyperparameters are the subset size and the maximum number of iterations. The subset size must be sufficient to accurately estimate the model parameters, while the number of iterations affects the likelihood of identifying the optimal subset that minimizes the median residual. An insufficient number of iterations may lead to suboptimal solutions, especially when the proportion of outliers is high, although increasing the iterations significantly raises the computational burden.

In the case of RANSAC, the hyperparameters include subset size, maximum iterations, residual threshold, and minimum number of inliers. Among these, the residual threshold is the most critical parameter as it determines which data points are classified as inliers or outliers. If the threshold is set too low, some valid inliers may be discarded, reducing model accuracy; conversely, if set too high, outliers may be incorrectly included as inliers, adversely affecting the estimation. Additionally, the number of iterations must be carefully chosen to balance the confidence in finding a good model against computational efficiency.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

For the S-Estimator, the tuning constant, maximum number of iterations, and initial scale estimate are the parameters that govern robustness and convergence characteristics. The tuning constant controls the trade-off between outlier robustness and statistical efficiency: smaller values increase robustness but reduce efficiency, while larger values improve efficiency at the expense of robustness. A good initial scale estimate is essential to ensure stable and rapid convergence. Poor initialization can result in slow convergence or suboptimal outcomes. The maximum number of iterations limits computational effort but must be sufficient to allow convergence under varying data conditions.

Table 3. Summary of Method Comparation

Aspect	LMedS	RANSAC	S-Estimator
Objective	Minimize the median of squared residuals	Maximize inliers by fitting models to random subsets	Minimize a robust estimate of the scale of residuals
Outlier Robustness	High	High (depends on threshold)	High
Computational Complexity	High; requires many iterations over all possible subsets (combinatorial in nature)	Medium; random sampling reduces complexity but needs many iterations for reliability	High; due to iterative scale and residual estimation
Efficiency	Lower than others	Good if tuned properly	High statistical efficiency when well-tuned
Breakdown Point (BP)	50%	≥ 50%	50%
Implementation Complexity	Conceptually straightforward, direct implementation can be costly.	Relatively simple and widely implemented.	More intricate, requiring algorithms for robust scale estimation and optimization.
Key Advantages	High breakdown point without the need for an explicit outlier threshold	Simplicity, effectiveness for moderate outlier rates, and explicit outlier identification.	High robustness to both outliers and leverage points, with good statistical efficiency.
Principal Limitations	Potential for significant computational cost.	Requires careful selection of the inlier threshold and the number of iterations.	Increased computational complexity and the need for judicious parameter selection.
Typical Applications	Robust regression in the presence of substantial outlier contamination.	Computer vision (feature matching, motion estimation), and robotics.	Robust regression analysis, particularly in datasets with potential high-leverage points

The Least Median of Squares (LMedS) method focuses on minimizing the median of squared residuals, making it highly effective at reducing the influence of extreme outliers and offering robustness up to nearly 50% data contamination. However, LMedS involves a combinatorial search process requiring numerous iterations, resulting in high computational complexity and inefficiency for large datasets or real-time applications. In contrast, the S-Estimator balances robustness and statistical efficiency by minimizing a robust scale of residuals. While it provides accurate estimates and performs well on clean data, it is highly sensitive to tuning parameters and initialization, which can limit its practicality for embedded systems or real-time applications with constrained resources. Meanwhile, the Random Sample Consensus (RANSAC) method employs random subset sampling and selects models based on the highest number of inliers. It demonstrates considerable robustness to outliers, with the advantage of greater computational efficiency compared to LMedS, and adapts well to various data conditions. However, RANSAC's performance strongly depends on proper residual threshold selection and the number of iterations.

4. CONCLUSION

The comparative analysis of Least Median of Squares (LMedS), Random Sample Consensus (RANSAC), and S-Estimators reveals that all three methods are robust techniques designed to reduce the influence of outliers in model estimation. LMedS excels in robustness without needing a predefined threshold but can be computationally intensive for large datasets. RANSAC offers a practical balance between robustness and efficiency through its iterative inlier maximization approach, though it requires careful parameter tuning. S-Estimators provide strong resistance to both outliers and high-leverage points backed by solid statistical theory, but their implementation is more complex. In the context of clock skew measurement, RANSAC is often preferred due to its effective outlier handling, ease of implementation, and explicit outlier identification, making it suitable for fast and repeated processing typical in distributed systems. However, LMedS and S-Estimators remain valuable for scenarios demanding higher robustness or more complex anomaly structures. Ultimately, method selection depends on dataset characteristics, anomaly types, and computational constraints. Empirical validation is essential to optimize parameters and ensure robust performance in practical applications..

REFERENCES

[1] Azmir, A. A., Wirastuti, N. M. A. E. D., & Saputra, K. O. (2022). Review methods for clock skew measurement . Matrix : Jurnal Manajemen Teknologi Dan Informatika, 12(3), 117–122. https://doi.org/10.31940/matrix.v12i3.117-122.

[2] Aditya, I. G. A. A. P., Saputra, K. O., & Wirastuti, N. M. A. E. D. (2022). Metode Potensial Untuk Menghitung Clock Skew. Journal of Computer Science and Informatics Engineering (J-Cosine), 6(2), 142-149.

[3] Putra Sastra, N., Oka Saputra, K., & Made Wiharta, D. (2021). Coexisting Parallelogram Method to Handle Jump Point on Hough Transform-based Clock Skew Measurement. Journal of Communications Software and Systems, 17(4), 297-304.

[4] Sastra, N. P., Saputra, K. O., & Teng, W. C. (2025). Offsets Reconstruction Method for Clock Skew Measurement Over High-Jitter Communication. SN Computer Science, 6(5), 404.

[5] Chen, K., Chen, X., & Gao, X. (2022, February). An anti-outlier robust clock compensated algorithm for wireless network. In Journal of Physics: Conference Series (Vol. 2189, No. 1, p. 012022). IOP Publishing.

[6] D. M. Khan, A. Yaqoob, S. Zubair, M. A. Khan, Z. Ahmad, and O. A. Alamri, "Applications of robust regression techniques: An econometric approach," Mathematical Problems in

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Engineering, vol. 2021, no. 1, p. 6525079, 2021.
[7] SETYOWATI, E., AKBARITA, R., & ROBBY, R. R. (2021). Perbandingan Regresi Robust Metode Least Trimmed Square (Lts) dan Metode Estimasi-S pada Produksi Padi di Kabupaten Blitar. Jurnal Matematika UNAND, 10(3), 329-341.
[8] De Brabanter, K., & De Brabanter, J. (2021). Robustness by Reweighting for Kernel Estimators: An Overview. Statistical Science, 36(4), 578-594.
[9] Cai, Q., Li, X., & Wu, Y. (2024). Linear Relative Pose Estimation Founded on Pose-only Imaging Geometry. arXiv preprint arXiv:2401.13357.
[10] C. Li, X. Li, T. Li, Q. Meng and W. Yu, "LMedS-Based Power Regression: An Optimal and Automatic Method of Radiometric Intercalibration for DMSP-OLS NTL Imagery," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2046-2057, 2021, doi: 10.1109/JSTARS.2021.3051800
[11] Mulyawan, B., Jovianto, N., Hendryli, J., & Herwindiati, D. E. (2020, July). Land mapping with least median of squares regression using landsat imagery: a case study Jakarta and sorrounding area. In IOP Conference Series: Materials Science and Engineering (Vol. 852, No. 1, p. 012024). IOP Publishing.
[12] Ratri, A. P., Susanti, Y., & Slamet, I. (2021). The Factors Affecting Soybean Production in Indonesia Using Robust Regression with Least Median of Squares (LMS) Estimation. Nusantara Science and Technology Proceedings, 70-78.
[13] Ren,Y. (2024). Performance Comparison of RANSAC and Other Model Estimation Methods in Panoramic Image Mosaic. Applied and Computational Engineering,105,82-90.
[14] Stanković, L., Brajović, M., Stanković, I., Lerga, J., & Daković, M. (2021). RANSAC-based signal denoising using compressive sensing. Circuits, systems, and signal processing, 40, 3907-3928.
[15] Elashry, A., Sluis, B., & Toth, C. (2021). Improving ransac feature matching based on geometric relation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 321-327.
[16] Zheng, J., Peng, W., Wang, Y., & Zhai, B. (2021). Accelerated RANSAC for accurate image registration in aerial video surveillance. IEEE Access, 9, 36775-36790.
[17] S. Y. Cao, R. Zhang, L. Luo, B. Yu, Z. Sheng, J. Li, and H. L. Shen, "Recurrent homography estimation using homography-guided image warping and focus transformer," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2023, pp. 9833–9842
[18] Rousseeuw, P., and V. Yohai. 1984. Robust regression by means of S-estimators. In Robust and nonlinear time series analysis, ed. J. Franke, W. Hardle, and D. Martin, Lecture notes in statistics, vol. 26, 256–72. Berlin and New York: Springer Verlag.
[19] Kehinde, A. O. O., Adesanya, K. K., & Onafowokan, M. A. (2023). Performance of the ordinary least squares estimator method of estimating regression parameters and some robust regression methods. International Journal Of Health Records & Information Management (Ijhrim), 6(1).
[20] A. Fitrianto and S. H. Xin, "Comparisons between robust regression approaches in the presence of outliers and high leverage points," BAREKENG: J. Ilmu Mat. dan Terapan, vol. 16, no. 1, pp. 243–252, 2022.
[21] Djara, V. A. D., Andriyana, Y., & Noviyanti, L. (2022). Modelling the prevalence of stunting toddlers using spatial autoregressive with instrument variable and S-estimator. Commun. Math. Biol. Neurosci., 2022.
[22] Fadhlan, M. F., & Sensuse, D. I. (2022). Knowledge Repository Design to Improve Knowledge

Management Process Capabilities: A Systematic Literature Review. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 6(2), 246-251.

[23] Alyosef, A. A., Elias, C., & Nürnberger, A. (2021, January). Localization and transformation reconstruction of image regions: An extended congruent triangles approach. In 2020 25th International Conference on Pattern Recognition (ICPR) (pp. 3884-3891). IEEE.

[24] Dewayanti, A. A. (2021). Robust Estimation in Regression Model for Handling Outlier and Heteroscedasticity. Jurnal Matematika Thales, 3(1).

[25] Lee, S. H., & Civera, J. (2022). RODIAN: Robustified Median. arXiv preprint arXiv:2206.02570.

[26] Khotimah, K., Sadik, K., & Kurnia, A. (2021, March). Robust multi-stage method (MM) and least median square (LMS) evaluation on handling outlier for multiple regression. In Journal of Physics: Conference Series (Vol. 1863, No. 1, p. 012033). IOP Publishing.

[27] Veríssimo, A., Lopes, M.B., Carrasquinha, E., Vinga, S. (2020). Random Sample Consensus for the Robust Identification of Outliers in Cancer Data. In: Cazzaniga, P., Besozzi, D., Merelli, I., Manzoni, L. (eds) Computational Intelligence Methods for Bioinformatics and Biostatistics. CIBB 2019. Lecture Notes in Computer Science, vol 12313. Springer, Cham. https://doi.org/10.1007/978-3-030-63061-4_11.

[28] Ortiz, A., Ortiz, E., Miñana, J.J., Valero, Ó. (2021). On the Use of Fuzzy Metrics for Robust Model Estimation: A RANSAC-Based Approach. In: Rojas, I., Joya, G., Català, A. (eds) Advances in Computational Intelligence. IWANN 2021. Lecture Notes in Computer Science(), vol 12861. Springer, Cham. https://doi.org/10.1007/978-3-030-85030-2_14.

[29] TA, S. (2023). Comparative Study on Robust Estimators and Evaluating their Performance in Multiple Regression. Aligarh Journal of Statistics, 43.

[30] Rahayu, D. A., Nursholihah, U. F., Suryaputra, G., & Surono, S. (2023). Comparasion of the m, mm and s estimator in robust regression analysis on indonesian literacy index data 2018. EKSAKTA: Journal of Sciences and Data Analysis, 11-22.

[31] Hakan Savaş Sazak & Nalan Mutlu (2023) Comparison of the robust methods in the general linear regression model, Communications in Statistics - Simulation and Computation, 52:7, 3163-3182, DOI: 10.1080/03610918.2021.1928196.

[32] S. Pradhan, S. Kundu, S. K. Ghosh, S. Chakraborty, M. Alam, S. S. Thakur, and B. K. S. Roy, "S-Estimator-Based Linear Robust Static State Estimation of Power Systems Considering Uncertain Noise Characteristics," in Proc. Int. Symp. Sustainable Energy and Technological Advancements, Singapore, Feb. 2024, pp. 243–261, Springer Nature Singapore.

[33] G. Canavire-Bacarreza, L. Castro Peñarrieta, and D. Ugarte Ontiveros, "Outliers in semi-parametric estimation of treatment effects," Econometrics, vol. 9, no. 2, p. 19, 2021.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id