Iraqi Journal of Statistical Sciences

https://stats.uomosul.edu.iq/

Integrating Wavelet Shrinkage with SURE and Minimax Thresholding to Enhance Maximum Likelihood Estimation for Gamma-Distributed Data

Hutheyfa Hazem Taha[1] Taha Hussein Ali [2] Heyam A. A. Hayawi[3]

[1] Department of Operations Research and Intelligent Techniques, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq.
[2] Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University, Erbil, Iraq.
[3] Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq.

Article information

Article history:
Received: April 1,2025
Revised: July 25,2025
Accepted: August 15,2025
Available online: December 1,2025

Keywords:
Gamma distribution, Maximum Likelihood Estimation, Wavelet Shrinkage, Data Contamination, and Parameter Estimation.

Correspondence:
Taha Hussein Ali
taha.ali@su.edu.krd
hutheyfa71@uomosul.edu.iq
he.hayawi@uomosul.edu.iq

Abstract

This paper uses the Maximum Likelihood Estimation method to investigate the impact of data contamination on the accuracy of parameter estimation for the Gamma distribution. A de-noising approach based on wavelet shrinkage has been proposed to address the limitations posed by contamination. Several types of wavelet functions were employed in combination with different threshold estimation techniques, namely Universal, Minimax, and Stein's Unbiased Risk Estimate, applying the soft thresholding rule. The study involved simulating data sets generated from the Gamma distribution and analyzing real-life data assumed to follow the same distribution. A specialized program was developed in MATLAB to conduct these simulations and implement both the classical Maximum Likelihood Estimation method and the proposed wavelet-based de-noising techniques. The performance of the parameter estimates was compared using the Mean Squared Error criterion. The findings demonstrated that data contamination significantly affects the accuracy of parameter estimates obtained through the classical Maximum Likelihood Estimation method. In contrast, the proposed wavelet shrinkage method effectively reduced the influence of contamination and enhanced the accuracy of parameter estimation for the Gamma distribution. The study highlights the practical value of integrating wavelet-based denoising techniques into statistical estimation processes, particularly when working with contaminated datasets.

DOI:10.33899/iqjoss.v22i2.54079,

1. Introduction

In many statistical applications, accurately estimating the parameters of probability distributions is fundamental for making reliable decisions, and among such distributions, the gamma distribution is of particular importance due to its wide use in such areas as probabilistic modelling of failure Times, analysis of environmental and medical data. However, the process of estimating the parameters of this distribution becomes complicated and sensitive in the presence of contaminated data or impurities (outliers), as these abnormal values lead to the deviation of estimates away from their true values, which impairs the efficiency and accuracy of statistical analysis (Zhao & Wang, 2012).

To overcome this problem, wavelet Shrinkage techniques have emerged as an effective tool for data processing and extracting the real signals underlying noise or pollution (Kerkhof & Molenaar, 2008). These techniques rely on converting data to the waveform domain, then applying shrinkage strategies to the waveform coefficients to remove or mitigate the effect of unwanted components (such as noise or outliers) and then reconstructing the purified signal. This method is particularly suitable for data that are heterogeneous or contain topical features that are difficult to detect by conventional means.

This research focuses on studying the effect of pollution in estimating the gamma distribution parameters and reviews the effectiveness of wave reduction techniques in mitigating these effects by analyzing contaminated data and comparing the estimation results before and after using wavelet processing. The research also seeks to identify the most appropriate minimization techniques (such as fixed or soft threshold minimization) and the best waveform bases that improve the accuracy of estimates, thereby enhancing the efficiency of the statistical models used (Elias and Ali, 2025).

2. Parameter Estimation for Gamma Distribution

The gamma distribution is one of the important continuous probability distributions, and it is widely used in statistical modelling, especially in areas related to positive and non-negative phenomena, such as waiting times, device lifetimes, water flows, and others.

The gamma distribution is defined by two basic parameters (Xiao Ke et al., 2023):

1. The shape parameter is often denoted by α
2. Measurement parameter denoted by θ

The probability density function of the gamma distribution takes the following form (Ozancan, 2021):

$$f(x;\alpha,\theta)=\frac{x^{\alpha-1}e^{\frac{-x}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)} \quad , \quad x>0 \tag{1}$$

Used to maximize the likelihood Function of the given data (MLE). Estimation equations are often non-linear and require numerical techniques (iterative algorithms such as the Newton-Raphson algorithm). This method is efficient under ideal conditions, but it is susceptible to contamination or outliers (Zhou, 2024).

Two gamma parameters are estimated by:

$$\hat{\alpha}: Solution\ to\ Ln(\hat{\propto})-\varphi(\hat{\alpha})=Ln(\overline{x})-\frac{1}{n}\sum_{i=1}^{n}Ln(x_i) \quad and \quad \hat{\theta}=\frac{\overline{x}}{\hat{\propto}} \tag{2}$$

Where: $\overline{x}$ is the average, and the digamma function $\varphi(\hat{\alpha})$ is the first derivative of the logarithm of the Gamma function (Loaiy and Huda, 2021).

3. Impact of Data Contamination on Statistical Estimation

In statistical analysis, it is often assumed that the collected data accurately represents the studied society. However, this assumption may not always be fulfilled, especially when data is contaminated, indicating the presence of anomalous values (Outliers) or atypical sightings that do not follow the same probability distribution as the data is assumed to have (Botani et al., 2025).

Causes of data contamination can stem from several factors, including errors in measurement or recording, merging data from different heterogeneous sources, rare or exceptional cases within the sample, and overlapping multiple distributions in one sample.

Types of pollution: Mild contamination refers to values that diverge slightly from the overall trend data, and severe contamination (Heavy Contamination) is characterized by data points that are substantially distant from the other values and may drastically alter the results of the analysis (Elias and Ali, 2025).

Structured contamination occurs when the contaminated values follow a specific pattern (e.g., they derive from a different distribution). The presence of contamination in the data can lead to deviations in estimates (bias), meaning that the average estimate does not reflect the true value of the parameter. Increased variability results in a lack of accuracy and instability in estimates. Loss of consistency arises when estimates fail to approach the true value as the sample size increases. There is a negative impact on hypothesis tests, such as a high error rate for the first or second type. Additionally, distortion of the overall shape of the distribution complicates the application of traditional probabilistic models.

Data pollution is a major challenge in applied statistics, and it directly affects the accuracy of estimation and analysis. It is therefore necessary to use pollution-resistant tools and methods when analyzing data, especially in cases where it is difficult to avoid or detect pollution directly.

4. Wavelet Shrinkage Techniques in Data Denoising

Noise in data is a common problem in many applied fields such as signal processing, image analysis, time series analysis, and experimental measurements. Noise refers to undesirable components that mix with the real signal and impair the accuracy of the analysis or statistical estimate.

One of the most effective methods of noise removal or mitigation is the wavelet Shrinkage technique, which has proven its efficiency in extracting useful signals from noise-contaminated data without losing fine details.

Wavelets are mathematical functions used to analyze signals and data at several levels of accuracy. Unlike conventional transformations (such as the Fourier transform), waveforms provide a positional analysis of both time and frequency, allowing the detection of sudden or localized changes in the signal.

The wavelet reduction technique is based on the following steps:

I. Transforming data to the wave domain (Wavelet Transform).
II. Transforming the original data into waveform coefficients that represent the signal at multiple levels.
III. Application of the shrinkage or threshing process.
IV. The wave coefficients with small values (thought to be caused by noise) are reduced or deleted, and the larger values (thought to represent the real signal) are retained.
V. Signal reconstruction (Inverse Wavelet Transform).
VI. Transform the purified coefficients back to the temporal or spatial domain to recreate the original signal without noise.

5. Evaluation Criteria

To measure the accuracy of the parameter estimated for the Gamma distribution for several samples (k), the mean squared error (MSE) can be used as in the following formula:

$$MSR(\hat{\theta}) = \frac{\sum_{i=1}^{k}\left(\theta_i - \hat{\theta}_i\right)^2}{k} \tag{3}$$

6. Proposed Method

The proposed method dealt with the outlier problem in data that has a Gamma distribution when its parameters are estimated (shape and scale) using the Maximum Likelihood Estimate method. The treatment is carried out through wavelet shrinkage, which includes:

I- Compute the Discrete Wavelet Transformation (DWT) coefficients for a wavelet $W(x)$, as Daubechies, Symlets, and Coiflets wavelets.

II- The threshold level δ is estimated by one of the methods (e.g., SURE, Minimax, and Universal threshold).

III- Thresholding rules, Soft is used to keep or kill the discrete wavelet coefficients. Thus, we get the modified DWT coefficients $MW(x)$.

IV- Compute the inverse of the modified DWT as in formula (4).

$$x^* = lnv\left(MW(x)\right) \tag{4}$$

V- Finally, the data x^*, The maximum likelihood estimators of a and b for the Gamma distribution are the solutions to the simultaneous equations (Johnson et al. 1994):

$$\log(\hat{a}) = log\left(\overline{x}^* \Big/ \prod_{i=1}^{n} x_i^*\right)^{1/n} \quad ; \quad a, b > 0 \quad ; \quad x \in (0, \infty) \tag{5}$$

$$\hat{b} = \overline{x}^*/\hat{a} \tag{6}$$

7. Simulation Design and Real Data Application

To show the effect of outliers on the maximum likelihood estimations of the shape and scale parameters for the gamma distribution and comparison between the classical and the proposed method in terms of efficiency and accuracy of the estimated parameters for the Gamma distribution, an experimental aspect was done by simulating the Gamma distribution, then an applied aspect of the real data based on the MSE criterion. And by designing a program in MATLAB (version 2020a) dedicated to this purpose.

7.1 Experimental Aspect

Six cases were selected for shape and scale parameters (1, 1) and (2, 0.5), respectively, the sample sizes (100, 200, and 300), and by adding (2, 3, and 4) outliers to the generated data from the gamma distribution. For the first experiment with $n = 100$, shown in Figure 1.

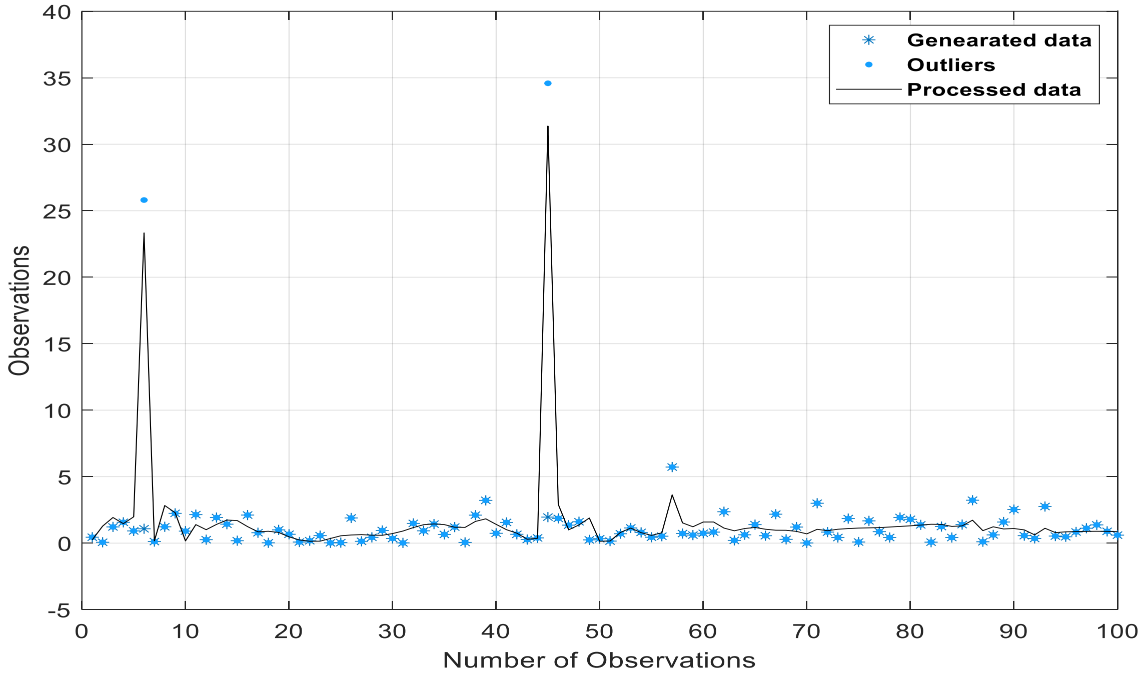

Figure (1): The generated data (*), Outliers (.), and processed data (-)

Figure 1 shows the scatter plot of the generated data from the Gamma distribution (*) at shape and scale parameters (1), and the two outliers (.), and then the data processed from the outliers (-) using the Daubechies wavelet and order of 5 (Db5) with Sure threshold and soft rule. The maximum likelihood estimators of a and b for the Gamma distribution, based on the data generated, which included outliers, and processed, are calculated in Table 1 along with the estimation error.

Table (1): MLE for two parameters and absolute estimation error

The data	$\hat{a}$	$\hat{b}$	Absolute estimation error (a)	Absolute estimation error (b)
generated data	1.1291	0.9141	0.1291	0.0859
which included outliers	0.6845	2.3457	0.3155	1.3457
Processed data	1.1570	1.4005	0.1570	0.4005

Table 1 shows the estimated parameter values (shape and scale) for the generated data. They were equal to (1.1291 and 0.9141) respectively, with absolute estimation error equal to (0.1291 and 0.0859) respectively, while the data added to outliers whose parameters reached (0.6845 and 2.3457) with a larger absolute estimation error equal to (0.3155 and 1.3457) respectively, due to the presence of the outliers. While the proposed method for processing data from outliers was the parameter values equal to (1.1570 and 1.4005), and the least absolute estimation error equal to (0.1570 and 0.4005), respectively.

Figures 2 and 3 show the probability density function and the cumulative probability function for the data generated from the gamma distribution with a sample size of 100, and the added data have outliers, and the process using the wavelet (Db5) and estimating the level of thresholding using SURE with the rule of soft thresholding.

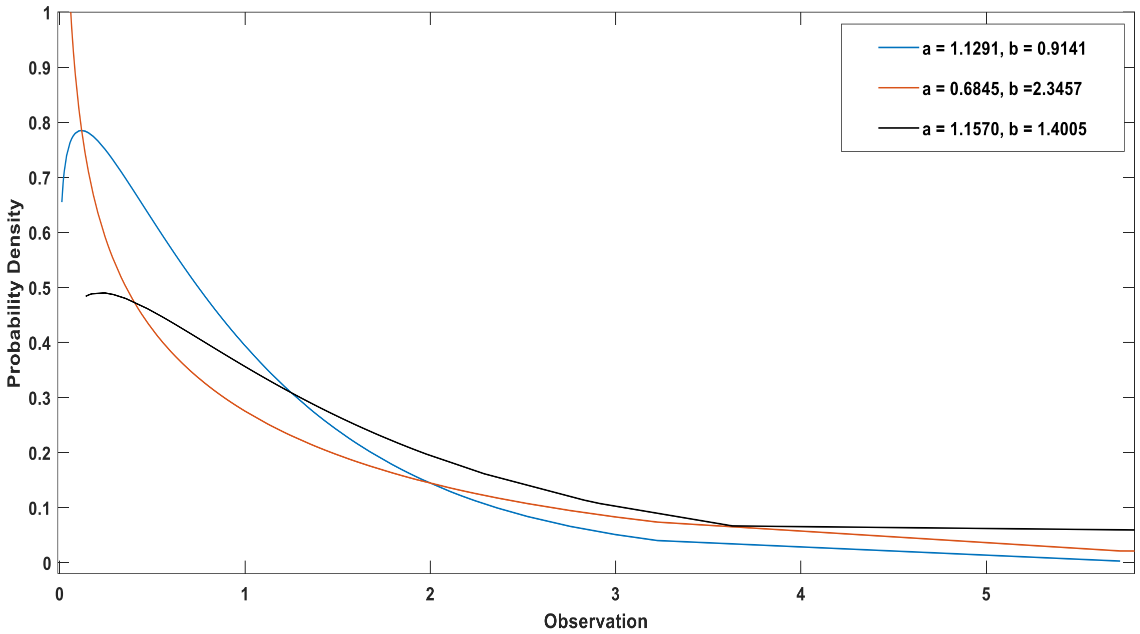

Figure (2): Gamma pdf for generated data (-), Outliers (-), and processed data (-)

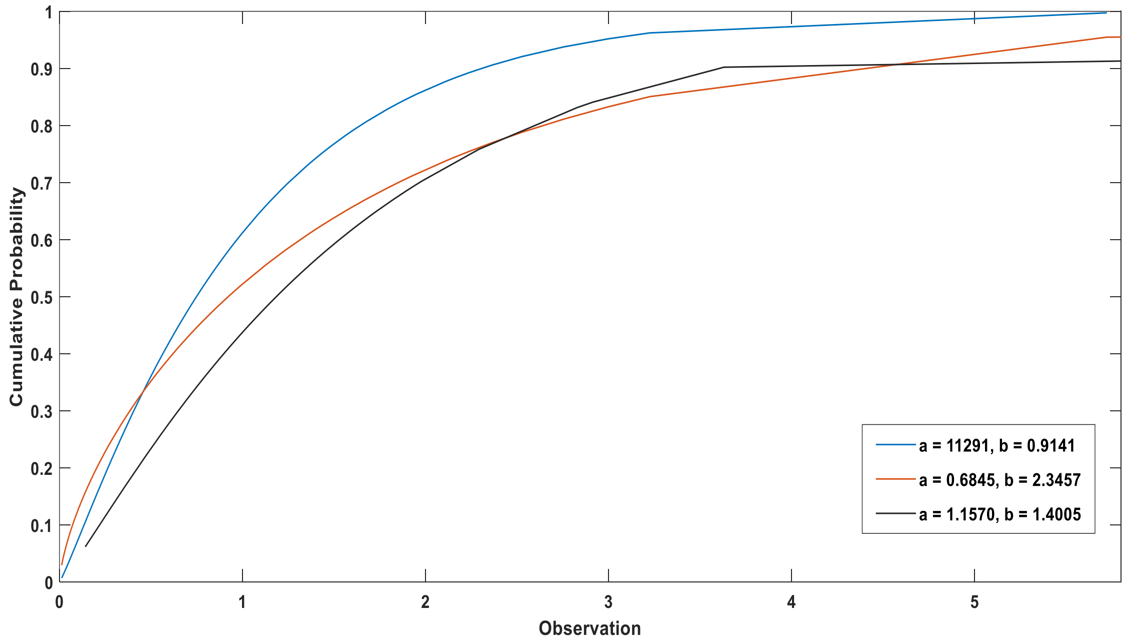

Figure (3): Gamma CDF for generated data (-), Outliers (-), and processed data (-)

For the comparison between the proposed and classical methods in estimating the parameter of the Gamma distribution, the experiment was repeated 1000 times, and the average criteria for MSE were calculated. Three wavelets (Db5), Symlets of order 1 (Sym1), and the Fejér–Korovkin wavelet of order 4 (Fk4) were used with different methods in estimating the threshold level (SURE, Minimax, and Universal), with threshold rule (Soft), and for different sizes (100, 200, and 300). The results are summarized in tables (2-7) for the average of the MSE criteria when at $x \sim Gamma(1, \quad 1)$, and $x \sim Gamma(2, \quad 0.5)$.

Table (2): Average of MSE Criteria When n = 100

Method	Shape and scale	Wavelet	Threshold	$AMSE(a)$	$AMSE(b)$
Proposed	(1, 1)	Db5	SURE	0.0536	2.6623
			Minimax	0.0551	1.9488
			Universal	0.1516	1.6278
		Sym1	SURE	0.0759	3.0802
			Minimax	0.0467	1.5543
			Universal	0.1112	**1.1675**
		Fk4	SURE	**0.0465**	2.2535
			Minimax	0.0532	1.6080
			Universal	0.1384	1.1832
Classical				0.1800	6.2428
Without outliers				0.0197	0.0258

Table (3): Average of MSE Criteria When n = 200

Method	Shape and Scale	Wavelet	Threshold	$AMSE(a)$	$AMSE(b)$
Proposed	(1, 1)	Db5	SURE	**0.0264**	1.2542
			Minimax	0.0386	0.8356
			Universal	0.0785	0.7688
		Sym1	SURE	0.0297	1.1711
			Minimax	0.0388	0.6703
			Universal	0.0865	**0.5116**
		Fk4	SURE	0.0276	1.1546
			Minimax	0.0379	0.7332
			Universal	0.0893	0.5420
Classical				0.1680	4.1465
Without outliers				0.0086	0.0127

Table (4): Average of MSE Criteria When n = 300

Method	Shape and scale	Wavelet	Threshold	$AMSE(a)$	$AMSE(b)$
Proposed	(1, 1)	Db5	SURE	**0.0160**	0.5356
			Minimax	0.0562	0.3040
			Universal	0.0866	0.3040
		Sym1	SURE	0.0207	0.4390
			Minimax	0.0721	0.2297
			Universal	0.1437	**0.1668**
		Fk4	SURE	0.0175	0.5060
			Minimax	0.0598	0.2613
			Universal	0.1398	0.1785
Classical				0.1442	2.6494
Without outliers				0.0058	0.0087

Table (5): Average of MSE Criteria When n = 100

Method	Shape and Scale	Wavelet	Threshold	$AMSE(a)$	$AMSE(b)$
Proposed	(2, 0.5)	Db5	SURE	1.4581	5.8050
			Minimax	1.2224	4.7752
			Universal	**0.9353**	3.8816
		Sym1	SURE	1.4417	5.6210
			Minimax	1.2345	4.1763
			Universal	1.1236	**3.6859**
		Fk4	SURE	1.3652	5.0663
			Minimax	1.2681	4.5217
			Universal	1.0880	3.7678
Classical				1.7246	7.7029
Without outliers				0.0896	0.0057

Table (6): Average of MSE Criteria When *n* = 200

Method	Shape and scale	Wavelet	Threshold	AMSE(a)	AMSE(b)
Proposed	(2, 0.5)	Db5	SURE	0.6541	0.8015
			Minimax	0.4699	0.6669
			Universal	0.3472	0.6055
		Sym1	SURE	0.5212	0.6255
			Minimax	0.3903	0.4915
			Universal	**0.3083**	**0.4187**
		Fk4	SURE	0.5682	0.6860
			Minimax	0.4317	0.5521
			Universal	0.3151	0.4521
Classical				1.1967	1.6730
Without outliers				0.0381	0.0028

Table (7): Average of MSE Criteria When *n* = 300

Method	Shape and scale	Wavelet	Threshold	AMSE(a)	AMSE(b)
Proposed	(2, 0.5)	Db5	SURE	0.2322	0.1985
			Minimax	0.1362	0.1427
			Universal	0.1361	0.1326
		Sym1	SURE	0.1577	0.1405
			Minimax	**0.1126**	0.0953
			Universal	0.1441	**0.0743**
		Fk4	SURE	0.1935	0.1645
			Minimax	0.1255	0.1137
			Universal	0.1352	0.0842
Classical				0.8915	0.7081
Without outliers				0.0249	0.0019

Tables 2–7 show that all the proposed methods have better efficiency than the classical method in estimating shape and scale parameters for the Gamma distribution, depending on the average of the criteria MSE for all cases. We also notice that there is a major impact on the accuracy of the estimated parameters due to the presence of outliers.

For comparison between the proposed methods, the best results for AMSE are summarized in Table 8. When x ~ Gamma (1, 1), the $AMSE(\alpha)$, the Fk4 wavelet was the best at a (100) sample size, while the Db5 wavelet was the best at (200 and 300) sample sizes with the SURE threshold method. When x ~ Gamma (2, 0.5), the $AMSE(\alpha)$, the Db5 wavelet was the best at a (100) sample size, while the Sym1 wavelet was the best at a (200 and 300) sample sizes with Universal threshold method at a (100 and 200) sample sizes and Minimax threshold method at a (300) sample size. The $AMSE(b)$**,** for the Sym1 wavelet with the Universal threshold method was the best to simulate all cases.

Table (8): The Best AMSE Criteria

Sample Size	Shape and Scale	Wavelet	Threshold Method	$AMSE(\alpha)$
100	(1, 1)	Fk4	SURE	0.0465
200	(1, 1)	Db5	SURE	0.0264
300	(1, 1)	Db5	SURE	0.0160
100	(2, 0.5)	Db5	Universal	0.9353
200	(2, 0.5)	Sym1	Universal	0.3083
300	(2, 0.5)	Sym1	Minimax	0.1126
Sample Size	**Shape and Scale**	**Wavelet**	**Threshold Method**	AMSE(b)
100	(1, 1)	Sym1	Universal	1.1675
200	(1, 1)	Sym1	Universal	0.5116
300	(1, 1)	Sym1	Universal	0.1668
100	(2, 0.5)	Sym1	Universal	3.6859
200	(2, 0.5)	Sym1	Universal	0.4187
300	(2, 0.5)	Sym1	Universal	0.0743

7.2 Application Part

The real data is taken from the source (Montgomery, 2012, p. 787). The sample size was 21 observations; the data were obtained from the operation of a factory for the transfer oxidation of ammonia to nitric acid for 21 days. The variable represents (stack loss X) with processed data from the outliers (-) using the (Sym3) wavelet with Universal threshold and soft rule, as shown in Figure 4.

The real and processed data about the gamma distribution were tested using the Kolmogorov-Smirnov (K. S.) and Chi-Square (χ^2) tests and are summarized in Table 9.

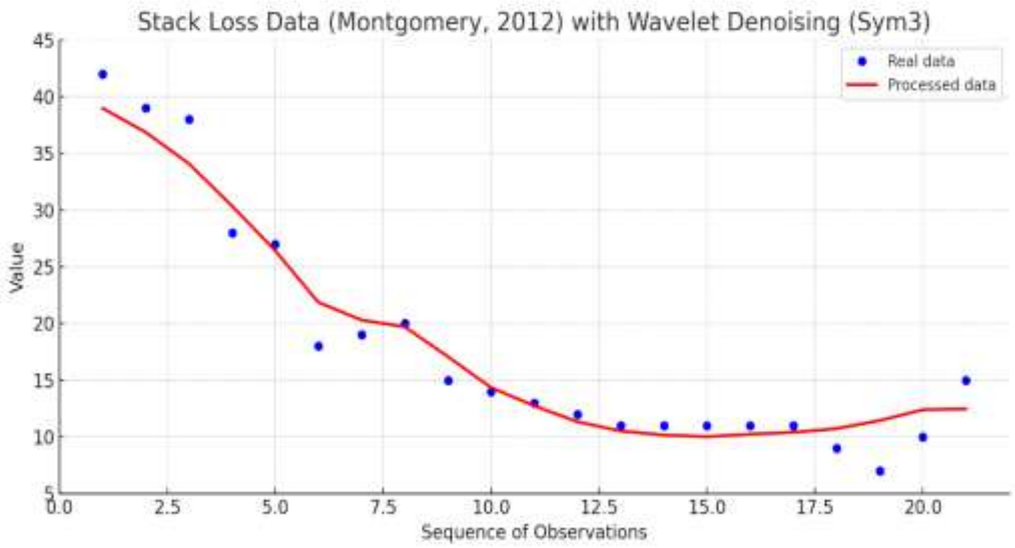

Figure (4): Real data (.), and processed data (-)

Table (9): Test of the Gamma distribution

The real data		Processed data	
K.S.	Chi-squared	K.S.	Chi-squared
0.14475 (0.71827)	3.365 (0.3387)	0.20228 (0.31295)	4.4692 (0.1070)
Critical values			
0.2872	7.8147	0.2872	5.9915

Table 9 shows that the (K.S.) and (χ^2) tests support the hypothesis that the real and processed data have a gamma distribution. Figure 5 for the box plot indicates the presence of three outliers in the real data, while it did not indicate their presence in the processed data. Outliers were deleted from the real data, using the box plot of the remaining data (18), which indicates that there are no outliers, as shown in Figure 5.

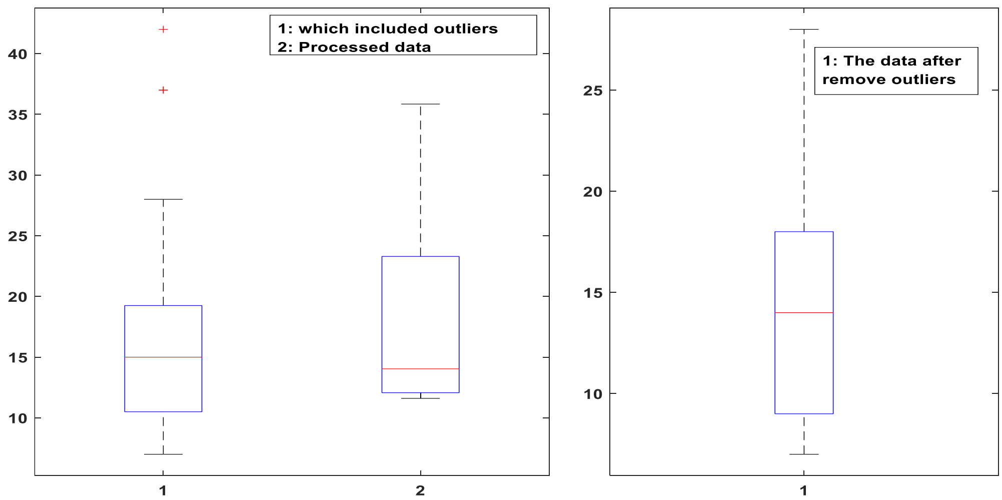

Figure (5): Box Plot for Real Data

Maximum likelihood estimates of a and b for the gamma distribution, based on real data that included outliers, data after outliers were removed, and processed with absolute error calculation, are summarized in Table 10.

Table (10): MLE for Two Parameters and Absolute Estimation Error

The data	$\hat{a}$	$\hat{b}$	Absolute Error (a)	Absolute Error (b)
The data after removing outliers	7.5493	1.8545	---------	---------
which included outliers	3.7910	4.6225	3.7583	2.7680
Processed data	6.6836	2.7535	0.8657	0.8990

Table 10 shows the estimated parameter values (shape and scale) for the data after removing outliers were equal to (7.5493 and 1.8545), respectively, while the data that included outliers whose parameters that reached (3.7910 and 4.6225), with a larger absolute estimation error equal to (3.7583 and 2.7680), respectively, due to the presence of the outliers. While the proposed method (using the (Sym3) wavelet with Universal threshold and soft rule) for processing data from outliers was the parameter values equal to (6.6836 and 2.7535), and the least absolute estimation error equal to (0.8657 and 0.8990), respectively.

Figure 6 shows the probability density function and the cumulative probability function for the real data (the same three cases).

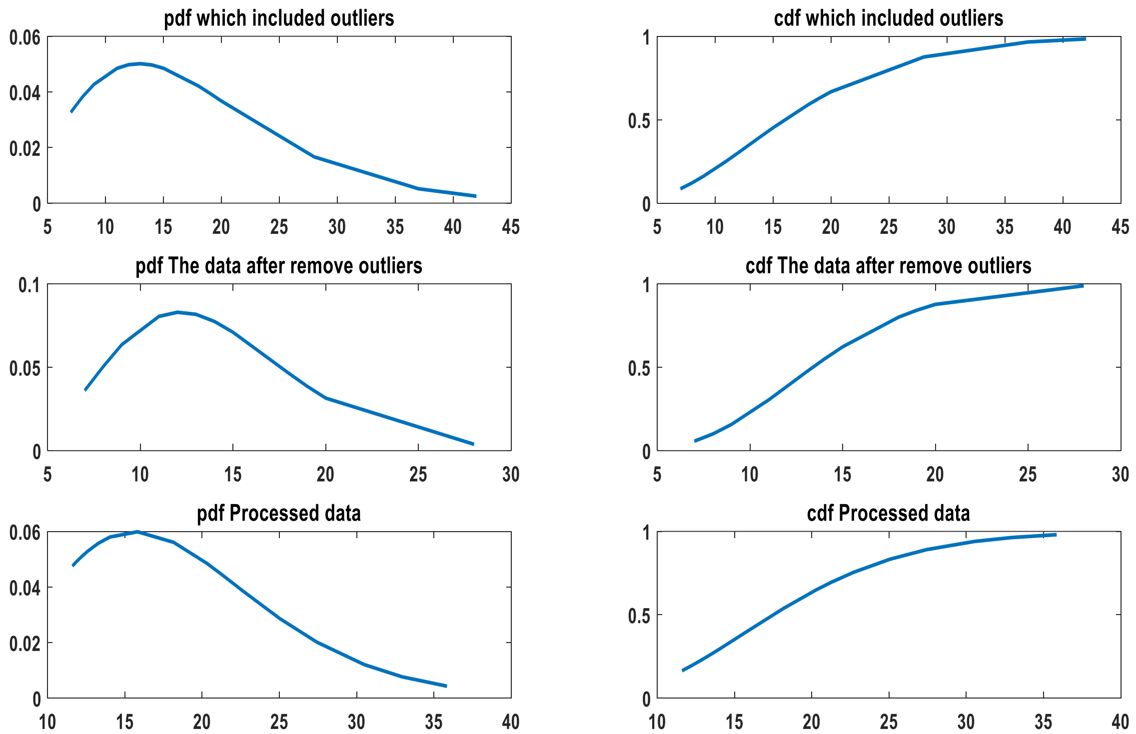

Figure (6): Gamma PDF and CDF for real data

In the first row of Figure 6 (Data After Removing Outliers), the PDF reveals a noticeable spread in the distribution with a peak at a moderate value and a long tail towards higher values. This indicates the presence of outliers, which skew the distribution and introduce asymmetry. The corresponding CDF rises gradually, reflecting the broader data spread. In alignment with the table, this scenario yielded a higher absolute error for both methods (3.7910 for method (a) and 4.6225 for method (b)), confirming that the presence of outliers negatively impacts estimation accuracy, particularly affecting method (b).

In the second row (data after removing outliers), the PDF becomes narrower and more symmetrical after removing outliers. The peak height increases, and the distribution's spread decreases, indicating that the data became more concentrated around the central values. The CDF curve also steepens, confirming the reduced variability. This reduction in data dispersion led to a noticeable improvement in absolute error, especially for method (b), which dropped to 1.8545, showing its sensitivity to outliers and substantial benefit from their removal. Method (a), however, saw its error increase to 7.5493, suggesting that it might rely on the extreme values removed or be more stable in the presence of outliers.

The third row (processed data) illustrates the PDFs and CDFs after applying data processing techniques. The PDF in this case becomes smoother and slightly broader than in Case 2, but without the heavy tails seen in Case 1. The CDF rises consistently without sudden jumps, reflecting a balanced distribution.

As shown in the table, both methods achieved their best performance in this case, with absolute errors significantly reduced (0.8657 for method (a) and 0.8990 for method (b)). This confirms that data processing not only minimizes the influence of outliers but also improves overall data structure, yielding more reliable estimation outcomes.

The figure clearly illustrates how the distribution characteristics- such as spread, symmetry, and concentration- evolve across the three cases. These visual patterns directly correspond to the changes in absolute errors for all cases; processed data consistently offers the most balanced distribution and optimal estimation results**.**

8. Conclusions

1. All the proposed methods have better efficiency than the classical method in estimating shape and scale parameters for the Gamma distribution, depending on the average of the criteria (MSE) for all cases.
2. There is a bad effect of outliers on the estimation quality of the gamma distribution parameters using MLE.
3. Sym1 wavelet with the Universal threshold method was the best to simulate all cases in estimating a scale parameter of the gamma distribution.
4. Fk4 wavelet was the best at a (100) sample size, while (Db5) wavelet was the best at a (200 and 300) sample sizes with the SURE threshold method at estimating a shape parameter (1) of the gamma distribution.
5. Db5 wavelet was the best at a (100) sample size, while (Sym1) wavelet was the best at a (200 and 300) sample sizes with Universal threshold method at a (100 and 200) sample sizes and Minimax threshold method at a (300) sample size to estimate a shape parameter (2) of the gamma distribution.
6. For real data, the proposed method for processing data from outliers and estimating the parameters of the gamma distribution was the best and had the lowest absolute estimation error.

References

1. Botani, D., Kareem, N., Ali, T., Sedeeq, B. (2025). Optimizing bandwidth parameter estimation for non-parametric regression using fixed-form threshold with Dmey and Coiflet wavelets. Hacettepe Journal of Mathematics and Statistics1-13. https://doi.org/10.15672/hujms.1605499
2. Elias, I. I., & Ali, T. H. (2025). Optimal level and order of the Coiflets wavelet in the VAR time series denoise analysis. Frontiers in Applied Mathematics and Statistics, 11, 1526540. https://doi.org/10.3389/fams.2025.1526540
3. Elias, I. Ibrahim, and Ali, T. Hussein. "VARMA Time Series Model Analysis Using Discrete Wavelet Transformation Coefficients for Coiflets Wavelet", Passer Journal of Basic and Applied Sciences, 7, 2, 2025, 657-677. doi: 10.24271/psr.2025.512121.2011. https://passer.garmian.edu.krd/article_225624.html.
4. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). Continuous Univariate Distributions, Volume 1 (2nd ed.). Wiley-Intercedence.
5. Kerkhof, G., & Molenaar, P. C. M. (2008). "Wavelet estimation in the presence of outliers." Computational Statistics & Data Analysis, 52(12), 5852–5865.
6. Loaiy F. Naji and Huda A. (2021). "Rasheed Bayesian Estimation for Two Parameters of Gamma Distribution under Generalized Weighted Loss Function", Ibn Al-Haitham Journal for Pure and Applied Sciences. **Link:** https://repository.uobaghdad.edu.iq/ articles/NheK8I4BVTCNdQwCF1q0re
7. Montgomery, D. C. (2012). Introduction to statistical quality control (7th ed., p. 787). Wiley. https://books.google.com/books/about/Statistical_Quality_Control_7th Edition. html?id=RgQcAAAAQBAJ
8. Ozancan Özdemir (2021)," Method of Moment Estimation for Two Parameter Gamma Distribution", DOI: 10.13140/RG.2.2.14110.31041.
9. Xiao Ke, Sirao Wang, Min Zhou, Huajun Ye (2023)." New Approaches on Parameter Estimation of the Gamma Distribution", Mathematics, 11, 927. **https://doi.org/10.3390/math11040927**
10. Zhao, Z., & Wang, D. (2012). "Wavelet threshold estimation for contaminated data." Statistics and Probability Letters, 82(3), 579–586.
11. Zhou Junmei, (2024)," Estimating parameters of the gamma distribution easily and efficiently", Communications in Statistics - Theory and Methods Volume 53, Issue 17, DOI: 10.1080/03610926.2023.2241097.

دمج انكماش الموجات مع Universal و Minimax و SURE لتحسين تقدير الإمكان الأعظم للبيانات الموزعة بواسطة جاما

حذيفة حازم طه[1] ، طه حسين علي[2] ،هيام عبد المجيد حياوي[3]

[1]قسم بحوث العمليات والتقنيات الذكائية، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق، [2]قسم الإحصاء والمعلوماتية، كلية الإدارة والاقتصاد، جامعة صلاح الدين، أربيل،.[3]قسم الإحصاء والمعلوماتية، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق.

الخلاصة: يستخدم هذا البحث طريقة تقدير الإمكان الأعظم للتحقيق في تأثير تلوث البيانات على دقة تقدير المعلمات لتوزيع جاما. تم اقتراح نهج لإزالة الضوضاء يعتمد على انكماش المويجات لمعالجة القيود التي يفرضها التلوث. تم استخدام عدة أنواع من دوال المويجات جنبا إلى جنب مع تقنيات تقدير العتبة المختلفة، وهي طرق تقدير Universal، Minimax ، وSURE غير المتحيز ، وتطبيق قاعدة العتبة الناعمة. تضمنت الدراسة محاكاة مجموعات البيانات الناتجة عن توزيع جاما وتحليل البيانات الحقيقية المفترضة لمتابعة نفس التوزيع. تم تطوير برنامج مخصص في ماتلاب لإجراء هذه المحاكاة وتنفيذ كل من طريقة تقدير الإمكان الأعظم الكلاسيكية وتقنيات إزالة الضوضاء القائمة على المويجات المقترحة. تمت مقارنة أداء تقديرات المعلمات باستخدام معيار متوسط الخطأ التربيعي. أظهرت النتائج أن تلوث البيانات يؤثر بشكل كبير على دقة تقديرات المعلمات التي تم الحصول عليها من خلال طريقة تقدير الإمكان الأعظم الكلاسيكية. في المقابل ، فإن طريقة انكماش المويجات المقترحة قللت بشكل فعال من تأثير التلوث وعززت دقة تقدير المعلمات لتوزيع جاما. تسلط الدراسة الضوء على القيمة العملية لدمج تقنيات تقليل الضوضاء القائمة على المويجات في عمليات التقدير الإحصائي, خاصة عند العمل مع مجموعات البيانات الملوثة.

الكلمات المفتاحية: توزيع كاما ، تقدير الإمكان الأعظم ، انكماش المويجات ، تلوث البيانات ، وتقدير المعلمات.