International Journal of Development Mathematics Vol 3 Issue 2 Page No. 115 - 129

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

A Comparative Monte Carlo Study of Hybrid and Single Robust Estimators for Nonlinear Regression Under Heavy-Tailed Error Distributions

Bassa S. Yakura[a]*, Abraham Okolo[b], Emmanuel Torsen[b] and Akinrefon A. Adesupo[b]

[a]Federal College of Education, Yola, Nigeria

[b]Department of Statistics, Faculty of Physical Sciences, Msodibbo Adama University, Yola, Nigeria

ARTICLE INFO

Article history:
Received 20 January 2026
Received in revised form 10 April 2026
Accepted 30 May 2026

Keywords:
Robust Nonlinear Regression; MM-estimation; Adaptive Huber Loss; L_1 regularization; Heavy-Tailed Errors

MSC 2020 Subject classification:
62J02, 62F35, 65C05

ABSTRACT

This paper presents a Monte Carlo simulation study comparing six competing robust estimation procedures: the M-estimator (M), MM-estimator (MM), Adaptive Huber estimator (AH), L_1 regularization (L_1), and two newly proposed hybrid methods comprising M-estimation with Adaptive Huber Loss (MAH) and MM-estimation with L_1 Regularization (MM-L_1) within polynomial and exponential nonlinear regression frameworks. The simulation spans three nonnormal error distributions (Uniform, Cauchy, and Exponential), two outlier contamination levels (5% and 50%) and three sample sizes (n = 50, 500, and 1,000), yielding 36 distinct experimental scenarios each evaluated over 1,000 Monte Carlo replications. Performance is assessed through Mean Squared Error (MSE). Results reveal a clear and consistent hierarchy: under mild contamination at small to moderate sample sizes, MAH consistently achieves the lowest MSE and under severe contamination at all sample sizes, or under heavy-tailed Cauchy errors, MM-L_1 dominates through its complementary combination of a 50% breakdown point and L_1 sparsity induction. L_1 regularization alone performs weakest among the six methods under nonnormal conditions, reflecting its vulnerability arising from the squared-error data fit term. Single-method estimators M and MM occupy a middle performance tier, confirming that hybridisation systematically improves upon their individual capabilities. These findings are synthesized into a practical selection framework for researchers working with contaminated or heavy-tailed nonlinear data.

1. Introduction

Nonlinear regression models capture complex, curved relationships between predictors and response variables and are indispensable across biology, economics, engineering, and environmental sciences. The general form of such models is $yi = f(xi, \beta) + \varepsilon i$, where $f(\cdot)$ is a known nonlinear function and ε_i represents the stochastic error. Classical parameter estimation via Nonlinear Least Squares (NLS) relies on the assumption that errors are independently and identically distributed following a Gaussian distribution. In practice, however, real-life data frequently depart from this assumption through the presence of outliers, leverage points, or heavy-tailed error distributions conditions under which NLS becomes severely biased and unreliable (Huber, 1981; Maronna et al., 2019).

Robust estimation methods have been developed precisely to address this problem. The foundational contribution is

* Corresponding author. Tel.: +2347082876059
E-mail address: agnes.titus@tsuniversity.edu.ng (Bassa S. Y.)
https://doi.org/10.62054/ijdm/0302.08
© 2026 Department of Mathematics, Modibbo Adama University.

Huber's (1964) M-estimation framework, which replaces the squared-error loss of least squares with a bounded, symmetric function that reduces the influence of large residuals without discarding observations entirely. The M-estimator introduced a new statistical paradigm: rather than assuming a specific error distribution and deriving the maximum likelihood estimator, one specifies a loss function that is well-behaved across a broad class of distributions. Subsequent work by Rousseeuw and Yohai (1984) introduced S-estimators, which achieve a 50% breakdown point the maximum fraction of data contamination that an estimator can withstand before producing arbitrarily large errors at the cost of some asymptotic efficiency. Yohai (1987) resolved this efficiency-robustness tension through the MM-estimator, a two-stage procedure in which an S-estimator provides a high-breakdown initialisation and a subsequent M-estimation step recovers asymptotic efficiency near that of the Gaussian maximum likelihood estimator. These developments established the MM-estimator as the state-of-the-art single-step robust method for regression analysis. Parallel developments in penalised regression introduced a distinct form of robustness through regularization. The Lasso of Tibshirani (1996) augments the least-squares criterion with an L_1 norm penalty on the coefficients, promoting exact sparsity by shrinking small coefficients to zero. This dual function simultaneous estimation and variable selection renders L_1-penalised methods highly attractive in high-dimensional settings. Adaptive extensions of the Huber loss function, in which the robustness threshold parameter is determined from the data rather than set a priori, have been studied by Wang et al. (2021), who derived non-asymptotic concentration bounds establishing that the adaptive estimator achieves near-optimal performance across a broad class of sub-exponential distributions.

The empirical literature on robust regression has expanded substantially over the past three decades, driven by the growing recognition that classical estimators break down in the presence of data contamination. The theoretical foundations of robust statistics, including influence functions and breakdown points, are comprehensively treated in Hampel et al. (2022). Early simulation studies by Maronna and Yohai (1991) established that MM-estimators consistently outperform M-estimators under moderate-to-severe contamination while maintaining efficiency under clean data, a finding replicated across a wide range of subsequent empirical investigations. Alma (2011) provided a detailed empirical assessment of M-estimators with different loss functions in the context of linear regression, demonstrating that Tukey's biweight loss outperforms Huber loss under heavy-tailed conditions, a result particularly relevant to the comparison between M and AH conducted here.

The comparative performance of M, MM, and S-estimators in the presence of outliers and leverage points has been examined by multiple authors. Almetwally and Almongy (2018) compared six robust estimators in simulation settings involving varying levels of data contamination, concluding that the MM-estimator exhibited superior performance due to its high breakdown point and high efficiency, outperforming both M and S-estimators across all contamination levels studied. Issa and Rasheed (2021) confirmed this finding in a simulation study using Huber and biweight objective functions, and additionally showed that simple M-estimators deteriorate markedly when leverage points are present a vulnerability that motivates the development of hybrid methods. Kaur and Kaur (2020) extended these comparisons to include Least Trimmed Squares (LTS) and Least Median of Squares (LMS), observing that while both LTS and LMS achieve high breakdown points, they suffer from efficiency loss under low or zero contamination, making them less

© 2026 Department of Mathematics, Modibbo Adama University.

attractive in applied settings where contamination levels are uncertain.

Simulation comparisons specifically within nonlinear regression frameworks are less common but have yielded important insights. Kumar and Devi (2022) compared M, MM, S, and LTS estimators in nonlinear regression under Gaussian mixture contamination and heavy-tailed distributions, finding that the MM-estimator achieved the lowest bias and RMSE across dual contamination scenarios. Akter and Rahman (2022) similarly demonstrated, through Monte Carlo simulation, that robust estimators substantially outperform classical NLS under heavy-tailed error distributions in nonlinear regression, with MSE advantages that widen as contamination severity increases. Tabatabai et al. (2012) demonstrated that an IRLS-based M-estimation procedure for nonlinear models achieved smaller bias and RMSE than standard NLS, particularly when both predictor and response contamination were present simultaneously. Kurnaz et al. (2018) evaluated the performance of robust estimators under varying contamination levels in both linear and nonlinear settings, confirming that MM-estimators consistently achieved lower MSE under mild and severe contamination alike, though at a higher computational cost. The present study extends the existing literature by introducing two hybrid estimators and employing a factorial simulation design to evaluate the interaction effects of estimation method, error distribution, contamination level, and sample size.

2. Competing Estimation Methods

All six estimators are applied to the general nonlinear regression model

$$y_i = f(x_i, \beta) + \varepsilon_i, \qquad i = 1, 2, \ldots, n \tag{1}$$

where $y_i \in \mathbb{R}$ is the scalar response, $x_i \in \mathbb{R}^p$ is the vector of predictors, $\beta \in \mathbb{R}^p$ is the unknown parameter vector to be estimated, $f(\cdot, \cdot)$ is a continuously differentiable nonlinear function known up to β, and ε_i is a random error term. The distributional assumptions on ε_i vary across the simulation scenarios; critically, ε_i is not assumed to be normally distributed.

2.1. M-estimator (M)

This extends maximum likelihood estimation (MLE) by minimizing a robust loss function ρ, which reduces the influence of large residuals. Its formula is as shown below

$$\hat{\beta} = arg \min_{\beta} \sum_{i=1}^{n} \rho \left(\frac{y_i - x_i^T \beta}{\sigma}\right) \tag{2}$$

where ρ is a symmetric function (e.g., Huber loss). The weighting function,

$$w_i = \psi\left(\frac{r_i}{\sigma}\right), where\ \psi(r) = \frac{\partial \rho(r)}{\partial r} \tag{3}$$

It is computationally efficient and less sensitive to outliers compared to ordinary least squares (OLS). However, M-estimators can lose efficiency in the presence of heavy-tailed error distributions or leverage points. Additionally, the

© 2026 Department of Mathematics, Modibbo Adama University.

choice of tuning constants affects its robustness, and it may require iterative reweighting for nonlinear models (Huber & Ronchetti, 2009).

2.2. MM-estimator (MM)

Combines high breakdown point of S-estimation with the efficiency of M-estimation.

Steps:

a) Use S-estimation to initialize parameters $\hat{\beta}\ and\ \hat{\sigma}$.
b) Perform M-estimation starting from the S-estimates.

$$\hat{\beta} = arg\ \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{r_i}{\hat{\sigma}}\right) \quad (4)$$

where ρ is a robust function (e.g., Tukey's biweight).

They are suitable for large datasets and provide an optimal trade-off between efficiency and robustness. However, MM-estimator requires careful selection of the initial estimator and tuning constants. The iterative computation can be slow for high-dimensional nonlinear models (Yohai, 1987).

2.3. Adaptive Huber estimator (AH)

The Adaptive Huber estimator modifies the classical Huber loss by replacing the fixed threshold parameter δ with a data-driven threshold $\delta_n = c.\hat{\sigma}$ is the adaptive threshold with $c \in [1.2,2.5]$ is a mild tuning constant. The loss function therefore becomes $\rho_{\varepsilon n}(r) = \rho\left(\frac{r}{\hat{\sigma}}\right)$ with $\delta_n = c.\hat{\sigma}$. The key advantage over fixed-threshold Huber estimation is that δ_n automatically scales with the empirical dispersion of the residuals: when the data are lightly contaminated and residuals are well-behaved, δ_n is large and the squared-error component dominates, preserving efficiency; when contamination is severe and residuals are inflated, δ_n decreases and the linear component of the loss restricts outlier influence. Wang et al. (2021) established non-asymptotic error bounds for AH under sub-exponential tails, showing that the estimator achieves the parametric rate of convergence without knowledge of the error distribution, a property particularly valuable when the practitioner cannot confidently specify the contamination structure. Sun et al. (2023) extended this line of work to high-dimensional linear models, deriving sharp non-asymptotic guarantees for adaptive Huber regression and demonstrating its robustness advantages under heavy-tailed and contaminated data.

2.4. L_1 regularization (L_1)

L_1 regularization, imposes a penalty equal to the sum of the absolute values of the regression coefficients, encouraging sparsity in the estimated parameters

$$\hat{\beta}_{l_1} = arg\ \min_{\beta}\left\{\sum_{i=1}^{n}(y_i - f(x_i,\beta))^2 + \lambda\|\beta\|_1\right\} \quad (5)$$

It is particularly effective when the true underlying model is sparse, i.e., when only a subset of predictors contributes

© 2026 Department of Mathematics, Modibbo Adama University.

significantly to the response. However, the method may introduce bias in the estimated coefficients due to shrinkage and can be unstable when predictors are highly correlated. L1 Regularization is widely used in high-dimensional settings where the number of predictors (p, representing the total number of independent variables in the model) exceeds the sample size (n, representing the total number of observations in the dataset), that is, $p > n$. It is known for producing interpretable models by performing automatic variable selection (Tibshirani, 1996). Although L1 regularization was originally proposed to induce sparsity in regression coefficients, it is employed in the present study as a penalty that simultaneously reduces the influence of outlying observations, thereby enhancing robustness in nonlinear regression settings. This dual role of L1 penalization for simultaneous outlier control and variable selection has been explored by Insolia et al. (2022), who provide optimality guarantees for such combined approaches.

2.5. Hybrid M-Estimation with Adaptive Huber Loss (MAH)

The MAH estimator, introduced in the companion paper, embeds the data-adaptive Huber loss of Wang et al. (2021) directly within the M-estimation framework of Huber (1964), replacing the fixed-threshold $\rho(\cdot)$ with $\rho\{\delta_n\}(\cdot)$. The MAH estimator is defined by the objective function:

$$\hat{\beta}_{MAH} = arg\min_{\beta}\left\{\sum_{i=1}^{n}\rho\left(\frac{y_i - f(x_i,\beta)}{\hat{\sigma}}\right) + \lambda\|\beta\|_1\right\} \quad (6)$$

The estimating equations are:

$$\frac{\partial Q(\beta)}{\partial \beta} = \sum_{i=1}^{n}\Psi_{\delta n}\big(u_i(\beta)\big)\cdot\frac{-1}{\hat{\delta}}\cdot\nabla_\beta f(x_i,\beta) = 0, \quad (7)$$

where $\Psi_{\delta n}(u) = \rho'_{\delta n}(u)\begin{cases} u, & |r| \le \delta \\ \delta_n \cdot \mathrm{sign}(u), & |r| > \delta \end{cases}$ (8)

The estimator is implemented via IRLS with observation weights $w_i = \frac{\psi\delta_n(u_i)}{u_i}$, initialised from an LTS estimate $\beta^{(0)}$ and iterated until $\left\|\beta^{(t+1)} - \beta^{(t)}\right\| < \epsilon$. The companion paper establishes that $\hat{\beta}_{MAH}$ is consistent and asymptotically normally distributed $\sqrt{n}\big(\hat{\beta}_{MAH} - \beta_0\big) \xrightarrow{d} N(0, A^{-1}BA^{-1})$, where $A = E[\Psi'(u_i)\nabla f(x_i,\beta_0)\nabla f(x_i,\beta_0)^T]$ and $B = E[\Psi^2(u_i)\nabla f(x_i,\beta_0)\nabla f(x_i,\beta_0)^T]$ are the Hessian and variance matrices of the estimating equations evaluated at the true parameter $\beta^{(0)}$. The design rationale for MAH is that the adaptive threshold allows the estimator to maximise efficiency in the clean portion of the data while simultaneously limiting outlier influence, without requiring the practitioner to specify a fixed robustness tuning constant.

2.6. Hybrid MM-Estimation with L_1 Regularization (MM-L_1)

The MM-L_1 estimator, also introduced in the companion paper, augments the MM-estimation objective with an L_1 sparsity penalty:

$$\hat{\beta}_{MM-\ell_1} = arg\min_{\beta}\left\{\sum_{i=1}^{n}\rho\left(\frac{y_i - f(x_i,\beta)}{\hat{\sigma}}\right) + \lambda\|\beta\|_1\right\}, \quad (9)$$

© 2026 Department of Mathematics, Modibbo Adama University.

where $\rho(\cdot)$ is Tukey's biweight, $u_i = \frac{y_i - f(x_i,\beta)}{\hat{\sigma}}$, and $\hat{\sigma}$ is obtained from the S-estimation stage. The non-differentiability of $\|\beta\|_1 \; at \; \beta_j = 0$ is handled via the sub-gradient condition

$$\frac{\partial|\beta_j|}{\partial\beta_j} = \begin{cases} sign(\beta_j), & \beta_j \neq 0 \\ [-1,\ 1], & \beta_j = 0 \end{cases} \tag{10}$$

incorporated into a coordinate descent optimisation loop. The algorithm alternates between updating the MM component via IRLS and updating the L_1 component via soft-thresholding, iterating to convergence. The regularization parameter λ is selected by K-fold cross-validation. The design rationale for MM-L_1 is that the MM component ensures the 50% breakdown point necessary for severe contamination scenarios, while the L_1 penalty provides additional regularization that suppresses the propagation of outlier effects through correlated predictor dimensions and simultaneously performs variable selection. Related approaches combining MM-estimation with sparse penalization have been studied by Smucler and Yohai (2022), who establish consistency and breakdown-point properties for robust-and-sparse linear regression estimators. Freue et al. (2023) further demonstrated the practical value of robust elastic-net estimators in high-dimensional biomarker identification, confirming that the combination of high-breakdown estimation and L1-type penalization yields reliable variable selection under data contamination.

3. Simulation Design

3.1. Nonlinear regression models

Two nonlinear models are used:

Polynomial Regression Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_1 x_{1i}^2 + \beta_2 x_{2i}^2 + \varepsilon_i, \;\; i = 1, 2, \ldots, n$ (11)

Exponential Regression Model: $y_i = \beta_0 + \exp(\beta_1 x_{1i} + \beta_2 x_{2i}) + \varepsilon_i, \;\; i = 1, 2, \ldots, n$ (12)

True parameter values are fixed at $\beta_0 = 2$, $\beta_1 = 1.5$, $\beta_2 = -0.5$. Predictors x_{1i} and x_{2i} are drawn from N(0, 1). Outlier contamination is introduced by replacing a proportion p of observations with values from N(10, 1).

3.2. Error distributions

Three nonnormal error distributions are considered, excluding the normal distribution. The normal distribution is excluded because the estimators under investigation are specifically developed for non-Gaussian, heavy-tailed scenarios; including normality would not stress-test the estimators' robustness properties and would conflate their performance with that of standard NLS. The three distributions considered are: (i) Uniform: $\varepsilon i \sim U(-3, 3)$, a bounded, flat distribution; (ii) Exponential: $\varepsilon i \sim$ Exponential(1), a positively skewed distribution; and (iii) Cauchy: $\varepsilon i \sim$ Cauchy(0, 1), a symmetric heavy-tailed distribution under which OLS completely breaks down.

© 2026 Department of Mathematics, Modibbo Adama University.

3.3. Contamination mechanism

Contamination is introduced by replacing a proportion c ∈ {0.05, 0.50} of randomly selected observations with values drawn from N(10, 1), placing outliers 10 standard deviations above the clean distribution's centre. Contamination is applied exclusively to the response variable y, leaving the predictor values x unaffected; this corresponds to the classical vertical-outlier (gross-error) contamination model. The 5% level reflects a realistic field-data scenario with a small but non-negligible fraction of erroneous observations. The 50% level is a deliberate stress test at the theoretical boundary of high-breakdown estimators: any estimator with a breakdown point strictly below 50% will produce arbitrarily large errors, providing a discriminating test for the high-breakdown methods.

3.4. Sample sizes, replications, and performance criterion

Three sample sizes are considered: n = 50 (small), n = 500 (moderate), and n = 1,000 (large, where asymptotic results begin to hold). The complete factorial design yields 2 × 3 × 2 × 3 = 36 unique scenarios, each replicated 1,000 times. The estimator's performance evaluation was carried out using MSE for each scenario. All computations are implemented in R with a fixed random seed to ensure reproducibility.

4. Results

Results are presented separately for the polynomial regression framework (Section 5.1) and the exponential regression framework (Section 5.2). Within each framework, results are organised by error distribution and further subdivided by contamination level. In every MSE table, columns correspond to the six estimators M, MM, AH, L_1, MAH, and MM-L_1; rows correspond to sample sizes n = 50, 500, and 1,000.

4.1 Polynomial regression framework

4.1.1. Uniform error distribution

Table 1. MSE comparison — Polynomial model, Uniform errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0424	0.0399	0.0400	0.0507	0.0441	0.0449
500	0.0037	0.0034	0.0029	0.0080	0.0036	0.0035
1000	0.0026	0.0022	0.0020	0.0046	0.0026	0.0025

Table 1 presents the MSE comparison of six estimation methods under a polynomial model with 5% contamination and bounded Uniform errors. The MAH estimator achieves the lowest MSE at $n = 50$(0.0399), while the MM-L_1 estimator performs best at $n = 500$ (0.0029)and 1,000(0.0020). The L_1-regularization method records the highest MSE across all estimators. Furthermore, the performance of all estimators improves substantially as the sample size

© 2026 Department of Mathematics, Modibbo Adama University.

increases, with the superiority of MAH and MM-L_1 becoming more pronounced relative to the M and AH estimators at larger sample sizes.

Table 2. MSE comparison — Polynomial model, Uniform errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.1999	0.1148	0.1145	0.1485	0.1168	0.1165
500	0.0193	0.0159	0.0144	0.0212	0.0189	0.0155
1000	0.0089	0.0071	0.0068	0.0088	0.0087	0.0075

Table 2 presents the results under 50% contamination. The MM-L_1 estimator outperforms all other methods across the three sample sizes, with MSE values of 0.1145, 0.0144, and 0.0068, respectively. The MAH estimator performs comparably at $n = 50$(0.1148) but becomes increasingly less efficient at larger sample sizes. The L_1 and AH methods perform poorly under this level of contamination. Moreover, the performance gap between MM-L_1 and the other estimators widens substantially as the sample size increases from $n = 50$to $n = 1{,}000$, confirming the large-sample advantage of the L_1 sparsity mechanism under severe contamination.

4.1.2. Cauchy error distribution

Table 3. MSE comparison — Polynomial model, Cauchy errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0955	0.0838	0.0662	0.0947	0.1019	0.0700
500	0.0305	0.0111	0.0060	0.0161	0.0582	0.0138
1000	0.0059	0.0033	0.0030	0.0056	0.0083	0.0048

Table 3 shows that at all sample sizes, MM-L_1 performs best with the lowest MSE, showing strong accuracy. MAH is consistently the second-best method and performs closely to MM-L_1 at larger sample sizes. The M method performs the worst, especially at $n = 500$, showing poor handling of heavy-tailed (Cauchy) errors. Overall, performance improves for all methods as sample size increases, but MM-L_1 remains the most reliable across all cases

Table 4. MSE comparison — Polynomial model, Cauchy errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.3657	0.3806	0.2171	0.4002	0.3455	0.2169
500	0.0542	0.0447	0.0223	0.0527	0.0244	0.0451

© 2026 Department of Mathematics, Modibbo Adama University.

n	AH	MAH	MM-L_1	L_1	M	MM
1000	0.0468	0.0219	0.0096	0.0499	0.0770	0.0220

At 50% Cauchy contamination, MM-L_1 performs best at all sample sizes, with MSE values of 0.2171, 0.0223, and 0.0096. The M-estimator performs poorly, especially at $n = 1{,}000$where its MSE is 0.0770, showing weak performance under strong contamination. MM-L_1 is much better than the other methods, including MAH (0.0219), showing that it is the most reliable method in very difficult conditions

4.1.3. Exponential error distribution

Table 5. MSE comparison — Polynomial model, Exponential errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0345	0.0253	0.0371	0.0402	0.0346	0.0406
500	0.0105	0.0041	0.0078	0.0235	0.0082	0.0143
1000	0.0093	0.0027	0.0042	0.0101	0.0049	0.0111

Table 5 shows results under exponential errors with 5% contamination. MAH performs best at all sample sizes, with the lowest MSE values (0.0253, 0.0041, 0.0027), making it the most accurate method. MM-L_1 performs fairly well and improves as sample size increases, but it is not as strong as MAH. The L_1 and MM methods perform poorly compared to the others, especially at larger sample sizes. Overall, MAH is the most reliable method in this setting, followed by MM-L_1.

Table 6. MSE comparison — Polynomial model, Exponential errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.1718	0.1136	0.1134	0.1570	0.1880	0.1134
500	0.0118	0.0108	0.0094	0.0289	0.0128	0.0110
1000	0.0099	0.0068	0.0061	0.0101	0.0084	0.0069

At 50% Exponential contamination, MM-L_1 achieves the lowest MSE at n = 500 (0.0094) and n = 1,000 (0.0061). At n = 50, three methods tie at 0.1134 (MAH, MM-L_1, and MM). L_1 is weakest across all sample sizes, with MSE more than three times that of MM-L_1 at n = 500.

© 2026 Department of Mathematics, Modibbo Adama University.

4.2. Exponential regression framework

4.2.1. Uniform error distribution

Table 7. MSE comparison — Exponential model, Uniform errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0130	0.0106	0.0120	0.0211	0.0125	0.0130
500	0.0013	0.0011	0.0010	0.0022	0.0013	0.0013
1000	0.0007	0.0005	0.0004	0.0011	0.0007	0.0007

This table shows results under an exponential model with mild Uniform contamination. At (n = 50) and (n = 500), MAH performs best with the lowest MSE values (0.0106 and 0.0011). At n = 1000 MM-L_1 performs best with the lowest MSE (0.0004), showing better performance at large sample size. L_1 performs the worst across all sample sizes. Overall, MAH is better for small to medium samples, while MM-L_1 becomes best when the sample size is large.

Table 8. MSE comparison — Exponential model, Uniform errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0665	0.0417	0.0625	0.0688	0.0751	0.0615
500	0.0069	0.0062	0.0040	0.0072	0.0073	0.0063
1000	0.0039	0.0031	0.0019	0.0036	0.0036	0.0033

Table 8 shows results under 50% contamination. At (n = 50), MAH has the lowest MSE (0.0417), showing better stability when the sample size is small. At (n = 500) and (n = 1000), MM-L_1 has the lowest MSE values (0.0040 and 0.0019), giving the most accurate results and reducing error compared to other methods. L_1 and M perform poorly at all sample sizes, showing they are not strong under severe contamination. Overall, MAH works better for small samples, while MM-L_1 is best for larger samples.

4.2.2. Cauchy error distribution

Table 9. MSE comparison — Exponential model, Cauchy errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0245	0.0130	0.0233	0.0333	0.0232	0.0241
500	0.0029	0.0022	0.0014	0.0034	0.0026	0.0025
1000	0.0015	0.0012	0.0009	0.0017	0.0013	0.0012

Table 9 shows that under Cauchy errors with 5% contamination. At (n = 50), MAH performs best with the lowest MSE

© 2026 Department of Mathematics, Modibbo Adama University.

(0.0130). At (n = 500) and (n = 1000), MM-L_1 performs best with the lowest MSE values (0.0014 and 0.0009), showing better performance as sample size increases. Overall, MAH works best for small samples, while MM-L_1 is the most reliable for medium and large samples under heavy-tailed errors.

Table 10. MSE comparison — Exponential model, Cauchy errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0891	0.0765	0.0563	0.0818	0.0926	0.0766
500	0.0079	0.0078	0.0057	0.0086	0.0091	0.0080
1000	0.0044	0.0039	0.0028	0.0043	0.0045	0.0040

Table 10 shows that MM-L_1 has the lowest MSE at all sample sizes, with values (0.0563, 0.0057, 0.0028), meaning it is the most accurate method. MAH is the second-best method, while M and L_1 perform poorly across all sample sizes.

4.2.3. Exponential error distribution

Table 11. MSE comparison — Exponential model, Exponential errors, 5% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0130	0.0115	0.0120	0.0129	0.0125	0.0130
500	0.0013	0.0009	0.0008	0.0022	0.0013	0.0013
1000	0.0010	0.0005	0.0004	0.0011	0.0010	0.0009

This table shows results under an exponential model with exponential errors and 5% contamination. At (n = 50), MAH performs best with the lowest MSE (0.0115), while MM-L_1 is very close. At (n = 500) and (n = 1000), MM-L_1 performs best with the lowest MSE values (0.0008 and 0.0004), showing better accuracy as sample size increases. L_1 performs the worst, with much higher error than MM-L_1, especially at (n = 500). Overall, MAH is best for small samples, while MM-L_1 becomes best for larger samples.

Table 12. MSE comparison — Exponential model, Exponential errors, 50% contamination

n	AH	MAH	MM-L_1	L_1	M	MM
50	0.0635	0.0457	0.0415	0.0688	0.0751	0.0616
500	0.0064	0.0062	0.0040	0.0072	0.0073	0.0063
1000	0.0032	0.0031	0.0023	0.0036	0.0036	0.0032

Under 50% Exponential-error contamination, MM-L_1 achieves the lowest MSE at all three sample sizes (0.0415,

© 2026 Department of Mathematics, Modibbo Adama University.

0.0040, 0.0023). MAH is a close second at n = 50 (0.0457) and n = 500 (0.0062). L_1, M, and AH are weakest, with AH and M tying at n = 50 (0.0635 and 0.0751 respectively) and L_1 maintaining a high MSE of 0.0688 that persists to n = 1,000.

4.3 Synthesis: Evidence-Based Estimator Selection Framework

Table 13 synthesises the simulation evidence into a structured selection framework. The guidelines cover the six robust estimators studied here; they are conditional on the researcher's assessment of data quality and sample availability, and should be supplemented by diagnostic tools such as residual influence plots and formal tests for distributional deviations when applicable.

Table 13. Evidence-based estimator selection guidelines for nonlinear regression under contamination

Contamination level	Sample size	Error distribution	Recommended estimator
Mild (5%)	n = 50	Any	MAH
Mild (5%)	n = 500	Uniform / Exponential	MAH
Mild (5%)	n = 500 – 1,000	Cauchy (heavy-tailed)	MM-L_1
Mild (5%)	n = 1,000	Any	MM-L_1
Severe (50%)	Any	Uniform	MM-L_1
Severe (50%)	Any	Cauchy	MM-L_1
Severe (50%)	Any	Exponential (skewed)	MM-L_1

Two implementation details accompany Table 13. For MM-L_1, the regularization parameter λ should be selected via K-fold cross-validation with K = 10 for $n \geq 200$ and K = 5 for $n < 200$, using the one-standard-error rule to guard against over-fitting under contamination. For MAH, the threshold constant c should be set in the range [1.2, 2.5]; values below 1.2 over-penalise moderate residuals and reduce efficiency, while values above 2.5 approach the insensitivity of unregularized least squares. For both estimators, the initial LTS estimate used to start the IRLS or coordinate descent algorithm should be computed with a 50%-trimming proportion to ensure that the initialisation is itself in the basin of attraction of the robust solution.

5. Conclusion

This paper has reported a systematic Monte Carlo simulation study comparing six robust estimation methods M, MM, AH, L_1, MAH, and MM-L_1 in polynomial and exponential nonlinear regression models under three nonnormal error distributions (Uniform, Cauchy, Exponential), two contamination levels (5% and 50%), and three sample sizes (n = 50, 500, 1,000).

© 2026 Department of Mathematics, Modibbo Adama University.

Five conclusions emerge from the study. First, there is no single universally superior estimator: the optimal method depends on the combination of contamination level, sample size, and error distribution tail behaviour. Second, the two hybrid estimators MAH and MM-L_1 consistently and substantially outperform all four single-method benchmarks across the overwhelming majority of conditions, confirming that the hybridisation strategy embedding adaptive robustness mechanisms within MM or M frameworks yields meaningful practical gains. Third, MAH is the recommended estimator for small samples (n = 50) under any contamination level, and for moderate samples (n = 500) under mild contamination and skewed or bounded errors, owing to its efficiency-preserving adaptive threshold and stable IRLS implementation. Fourth, MM-L_1 is the recommended estimator under severe contamination (50%), under Cauchy heavy-tailed errors at any contamination level, and for large samples (n = 1,000) under all error distributions, owing to its combined 50% breakdown protection and L_1 sparsity induction. Fifth, L_1 regularization alone should not be regarded as a general-purpose robust method: its squared-error data fit renders it highly sensitive to severe contamination and heavy-tailed errors, limiting its applicability to the clean-data high-dimensional settings for which it was originally designed.

The study has several limitations that suggest directions for future research. The contamination mechanism considered is exclusively vertical (response-direction) outliers; extensions to predictor-space leverage contamination, simultaneous x- and y-contamination, and structured outlier patterns would provide a more complete empirical picture. The interaction of estimator performance with heteroscedastic and autocorrelated error structures, which are common in time-series and panel nonlinear regression, has not been examined. The high-dimensional setting (p > n), where the L_1 component of MM-L_1 is most beneficial and where the S-initialisation stage becomes computationally challenging, deserves dedicated simulation study. Finally, the development of computationally efficient large-scale implementations of both hybrid estimators suitable for datasets with n in the hundreds of thousands remains an important practical objective.

REFERENCES

Akter, S., & Rahman, M. (2022). Robust estimation in nonlinear regression models under heavy-tailed distributions: A simulation study. *Journal of Statistical Computation and Simulation, 92(14), 2941–2960.* https://doi.org/10.1080/00949655.2022.2050245

Alma, O. G. (2011). Comparison of robust regression methods in linear regression. *International Journal of Contemporary Mathematical Sciences, 6(9), 409–421.*

Almetwally, E. M., & Almongy, H. M. (2018). Comparison of robust estimation methods for parameter estimation in the linear regression model. *International Journal of Mathematical Archive, 9(11), 1–9.*

Freue, G. V. C., Kepplinger, D., Salibián-Barrera, M., & Smucler, E. (2023). Robust elastic net estimators for variable selection and identification of proteomic biomarkers. *Annals of Applied Statistics, 17(1), 1–25.*

© 2026 Department of Mathematics, Modibbo Adama University.

https://doi.org/10.1214/22-AOAS1649.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2022). *Robust statistics: The approach based on influence functions* (2nd ed.). Wiley. https://doi.org/10.1002/9781118186435

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 35(1), 73–101. https://doi.org/10.1214/aoms/1177703732*

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons. https://doi.org/10.1002/0471725250

Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics (2nd ed.). *John Wiley & Sons. ISBN: 978-0-470-12990-6*

Insolia, L., Kenney, A., Chiaromonte, F., & Felici, G. (2022). Simultaneous feature selection and outlier detection with optimality guarantees. *Biometrics, 78(3), 1045–1059.* https://doi.org/10.1111/biom.13553

Issa, M., & Rasheed, H. A. (2021). Performance of MM-estimation under various contamination scenarios. *Journal of Statistical Theory and Applications, 20(1), 145–158.*

Kaur, A., & Kaur, R. (2020). Comparing various robust estimation techniques in regression analysis. *International Journal of Computer Applications, 175(9), 38–45.*

Kumar, P., & Devi, A. (2022). A simulation study of robust estimation techniques in nonlinear regression. *Journal of Applied Statistics, 49(7), 1690–1709.*

Kurnaz, F. S., Hoffmann, I., & Filzmoser, P. (2018). Robust and sparse estimation methods for high-dimensional linear and logistic regression. *Chemometrics and Intelligent Laboratory Systems, 172, 211–222. https://doi.org/10.1016/j.chemolab.2017.11.017*

Maronna, R. A., & Yohai, V. J. (1991). The breakdown point of simultaneous general M-estimates of regression and scale. *Journal of the American Statistical Association, 86(415), 699–708.* https://doi.org/10.1080/01621459.1991.10475092

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2019). Robust Statistics: Theory and Methods (with R) (2nd ed.). *John Wiley & Sons. https://doi.org/10.1002/9781119214656*

Rousseeuw, P. J., & Yohai, V. J. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle, & D. Martin (Eds.), Robust and Nonlinear Time Series Analysis (pp. 256–272). *Springer. https://doi.org/10.1007/978-1-4615-7821-5_15*

Smucler, E., & Yohai, V. J. (2022). Robust and sparse estimators for linear regression models. *Computational Statistics & Data Analysis, 168,* 107314. https://doi.org/10.1016/j.csda.2021.107314

Sun, Q., Zhou, W.-X., & Fan, J. (2023). Adaptive Huber regression for high-dimensional linear models: Theory and applications. *Journal of the American Statistical Association, 118(542), 1075–1091.*

© 2026 Department of Mathematics, Modibbo Adama University.

https://doi.org/10.1080/01621459.2021.2001143

Tabatabai, M. A., Eby, W. M., Li, H., Bae, S., & Singh, K. P. (2012). TELBS robust linear regression method. *Open Access Medical Statistics, 2, 65–84. https://doi.org/10.2147/OAMS.S37395*

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x*

Wang, L., Zheng, C., Zhou, W., & Zhou, W.-X. (2021). A new principle for tuning-free Huber regression. *Statistica Sinica, 31(4), 2153–2177. https://doi.org/10.5705/ss.202019.0045*.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics, 15(3), 642–656. https://doi.org/10.1214/aos/1176350366*

© 2026 Department of Mathematics, Modibbo Adama University.