METAHEURISTIC VERSUS CLASSICAL OPTIMIZATION FOR MAXIMUM LIKELIHOOD ESTIMATION IN LAMBERT–TOPP–LEONE REGRESSION MODELS

Azhar Hussein Alwan[1], Ahmed Mahdi Salih[2*]

[1]University of Diyala, Baqubah, Diyala Governorate, Iraq
[2]University of Wasit, Kut, Wasit Governorate, 52001, Iraq
*Corresponding author, email: amahdi@uowasit.edu.iq

doi: 10.17977/um066.v6.i9.2026.1

Keywords
Lambert–Topp–Leone distribution
Bounded regression
Maximum likelihood estimation
Grey wolf optimizer
Differential evolution
BFGS
Metaheuristic optimization

Abstract
This paper develops a Lambert-Topp-Leone (LTL) regression model for limited responses and explores the use of classical and metaheuristic optimisation strategies for maximum likelihood estimation. Grey Wolf Optimiser (GWO) and Differential Evolution (DE) are compared with BFGS and Nelder-Mead using Monte Carlo simulations with different parameter configurations and sample sizes. The performance is evaluated based on bias, RMSE, likelihood achievement, convergence dependability, and computing time. The results show that BFGS performs well under normal likelihood setups, while GWO and DE become competitive under more challenging parameter choices, showing that neither optimiser is uniformly superior.. An application to the MockJurors data indicates that the four techniques yield roughly the same LTL maximum likelihood solutions, lending support to the employment of metaheuristic algorithms as complementary global-search tools for assessing the numerical stability of LTL regression estimates.

1. Introduction

Regression models for bounded and proportion-valued responses have received considerable attention in statistical modeling because such data arise in many scientific, economic, biomedical, and engineering applications. Classical Gaussian regression is generally inappropriate when the response is restricted to the unit interval, since bounded observations commonly exhibit asymmetry, heteroscedasticity, and non-normal behavior. Beta regression provides a well-established framework for responses in (0, 1) (Cribari-Neto & Zeileis, 2010; Ferrari & Cribari-Neto, 2004), while several alternative distributions and generated families have been proposed to provide greater flexibility for bounded data (Bourguignon et al., 2014; Kieschnick & McCullough, 2003).

The Topp–Leone distribution, originally introduced by Topp and Leone (1955), is an important bounded distribution with a simple analytical structure. Its flexibility has motivated several extensions and generated families (Al-Shomrani et al., 2016; Rezaei et al., 2017). More recently, Astorga and Iriarte (2025) introduced the Lambert–Topp–Leone (LTL) distribution by incorporating a Lambert-type transformation into the Topp–Leone framework. The resulting model provides additional shape flexibility and has demonstrated potential for modeling proportion and lifetime data. This development also motivates the construction of regression structures in which LTL parameters are related to explanatory variables.

Maximum likelihood estimation (MLE) is a natural way to estimate the parameters of such regression models, due to its desired large-sample features and universal application (Casella & Berger, 2002; Pawitan, 2001). However, the likelihood functions of flexible nonlinear distributions may exhibit parameter limitations, high dependency between parameters, and irregular or relatively complicated objective surfaces. Numerical maximisation can be sensitive to beginning values and optimisation approach. Classical approaches such as BFGS (Nocedal & Wright, 2006) or the Nelder–Mead simplex algorithm (Nelder & Mead, 1965) are frequently used to optimise the

likelihood. However, they may perform poorly in the presence of challenging regions of the objective surface or in the absence of good initial values.

Metaheuristic optimization offers an alternative computational strategy by undertaking more thorough searches of the parameter space and generally requiring less reliance on derivatives or carefully selected initial solutions (Talbi, 2009; Yang, 2010). Mirjalili et al. (2014) introduced the Grey Wolf Optimizer (GWO), a population-based algorithm that draws inspiration from the foraging mechanism and leadership hierarchy of grey wolves. Storn and Price (1997) presented Differential Evolution (DE), a significant global optimization technique that utilises mutation, crossover, and selection to generate candidate solutions across continuous parameter domains. Differential Evolution (DE) has shown to be an efficient general-purpose optimizer owing to its very simple implementation and efficiency.

Driven by these factors, this research formulates a The Lambert-Topp-Leone regression framework is examined, focusing on the numerical characteristics of maximum likelihood estimation across several optimisation techniques. Specifically, the GWO and DE metaheuristic algorithms are juxtaposed with the traditional BFGS and Nelder-Mead techniques. The analysis is performed by comprehensive Monte Carlo simulations utilising metrics including bias, mean squared error, convergence characteristics, optimised log-likelihood, and computing duration. A practical application utilising real data is also evaluated to demonstrate the efficacy of the suggested regression model and to investigate the similarities and discrepancies among the optimisation methods.

In this way, this work makes two important contributions. First, it creates a regression framework using the new LTL distribution, which lets the features of the distribution change consistently with covariate data. Second, it gives a thorough look at both metaheuristic and classical optimisation methods for finding the best estimates of likelihood in the final model. We show that global and conventional optimisation approaches for LTL regression estimation are similar in terms of how accurate, stable, and time-efficient they are.

2. Method

2.1. The Lambert–Topp–Leone Regression Model

2.1.1. The Lambert–Topp–Leone distribution

Let Y be a continuous random variable defined on (0, 1). Following Astorga and Iriarte (2025), the Lambert–Topp–Leone (LTL) distribution provides a flexible extension of the classical Topp–Leone model (Topp & Leone, 1955). Its construction combines the bounded Topp–Leone structure with a Lambert-type transformation, thereby allowing additional control over the shape of the distribution.

For regression purposes, it is convenient to write the LTL density in the generic form

$$\mathbf{f(y;\,\theta,\lambda) = fLTL(y;\,\theta,\lambda),\quad 0 < y < 1,\ (1)}$$

where $\theta > 0$ is a distributional parameter and λ denotes the additional parameter associated with the Lambert transformation. The corresponding cumulative distribution function is denoted by

$$\mathbf{F(y;\,\theta,\lambda) = FLTL(y;\,\theta,\lambda).\quad (2)}$$

The parameter space is restricted so that $f(y;\,\theta,\lambda) > 0$ and $0 < F(y;\,\theta,\lambda) < 1$.

The additional flexibility of the LTL distribution is particularly useful for bounded responses exhibiting different degrees of skewness and concentration. This motivates extending the distribution to a regression setting in which one of its parameters varies systematically with explanatory variables.

2.1.2. Regression formulation

Consider independent observations $Y_1, \ldots, Y_n$, where $Y_i \in (0, 1)$, together with a vector of explanatory variables $x_i = (1, x_{i1}, \ldots, x_{ip})^T$. Assume that

$$\mathbf{Y_i \sim LTL(\theta_i, \lambda), \quad i = 1, \ldots, n.} \qquad \mathbf{(3)}$$

To ensure positivity of θ_i, we employ the logarithmic link

$$\mathbf{\log(\theta_i) = \eta_i = x_i^T \beta,} \quad \mathbf{(4)}$$

or equivalently,

$$\mathbf{\theta_i = \exp(x_i^T \beta),} \qquad \mathbf{(5)}$$

where
$\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$

is the vector of regression coefficients. The parameter λ is initially assumed to be common to all observations. This formulation allows the distribution of the bounded response to change with the covariates while preserving the required parameter restrictions. The sign and magnitude of each β_j describe the effect of the corresponding covariate on θ_i through the log link.

2.1.3. Likelihood inference

Let

$$\mathbf{\psi = (\beta^T, \lambda)^T}$$

denote the complete parameter vector. Under independence, the likelihood function is

$$\mathbf{L(\psi) = \prod_{i=1}^{n} fLTL(y_i; \exp(x_i^T \beta), \lambda),} \quad \mathbf{(6)}$$

and the corresponding log-likelihood is

$$\mathbf{\ell(\psi) = \sum_{i=1}^{n} \log fLTL(y_i; \exp(x_i^T \beta), \lambda).} \qquad \mathbf{(7)}$$

The maximum likelihood estimator is therefore defined as

$$\mathbf{\hat{\psi} = \arg\max_{\psi \in \Psi} \ell(\psi),} \qquad \mathbf{(8)}$$

where Ψ denotes the admissible parameter space.

Closed-form solutions are generally unavailable because the regression parameters enter the LTL likelihood nonlinearly. Consequently, numerical optimization is required. In this study, the maximization problem in (8) is solved using two classical procedures, BFGS and Nelder–Mead, and two population-based metaheuristic procedures, GWO and DE. The four algorithms therefore target the same maximum likelihood solution; their differences concern the numerical search strategy used to locate that solution.

2.2. Optimization Methods for Maximum Likelihood Estimation

Let ψ denote the vector containing the regression coefficients and the remaining LTL parameters. Estimation is based on maximizing the log-likelihood

$$\mathbf{\hat{\psi} = \arg\max_{\psi \in \Psi} \ell(\psi).} \qquad \mathbf{(9)}$$

Equivalently, the numerical algorithms minimize the objective function

$$\mathbf{Q(\psi) = -\ell(\psi).} \qquad \mathbf{(10)}$$

Four optimization procedures are considered: Grey Wolf Optimizer (GWO), Differential Evolution (DE), BFGS, and Nelder–Mead. The same likelihood, parameter space, and convergence criterion are used whenever applicable to ensure a fair comparison.

2.2.1. Grey Wolf Optimizer

The Grey Wolf Optimiser (GWO), which was created by Mirjalili et al. (2014), is a population-based metaheuristic that is based on how grey wolves hunt and run their packs. Possible answers can be broken down into four groups: alpha, beta, delta, and omega wolves. The first three were the best ones found during the search. The encircling mechanism for a possible answer X(t) is written as

$$\mathbf{D = |CX_p(t) - X(t)|, \quad X(t+1) = X_p(t) - AD, \quad (11)}$$

where

$$\mathbf{A = 2ar_1 - a, \quad C = 2r_2, \qquad (12)}$$

r_1 and r_2 contain independent uniform random numbers on (0, 1), and a decreases from 2 to 0 during the iterations.

The three best solutions, Xα, Xβ, and Xδ, guide the population. The updated position is

$$\mathbf{X(t+1) = (X_1 + X_2 + X_3) / 3. (13)}$$

In the present study, each wolf represents a possible value of ψ, and its fitness is evaluated using Q(ψ) in (10).

2.2.2. Differential Evolution

Differential Evolution (DE) is a stochastic population-based optimization method introduced by Storn and Price (1997). For each target vector $X_{i,t}$, a mutant vector is generated as

$$\mathbf{V_{i,t} = X_{r1,t} + F(X_{r2,t} - X_{r3,t}), \quad (14)}$$

where r_1, r_2, and r_3 are distinct population indices and $F \in (0, 2)$ is the mutation factor.

A trial vector is then produced through crossover between $V_{i,t}$ and $X_{i,t}$. If the trial vector gives a smaller value of the negative log-likelihood, it replaces the current solution:

$$\mathbf{X_{i,t+1} = \{ U_{i,t}, \ Q(U_{i,t}) \le Q(X_{i,t}); \ X_{i,t}, \ otherwise. \} \qquad (15)}$$

The procedure continues until the specified stopping condition is reached (Price et al., 2005).

2.2.3. Classical optimization methods

Two typical optimisation approaches are compared. BFGS is a quasi-Newton approach that iteratively approximates the inverse Hessian matrix and uses gradient information to optimise search direction (Nocedal & Wright, 2006). Rapid local convergence occurs when the goal function is smooth and the initial solution is appropriate.

The Nelder-Mead algorithm is a derivative-free simplex technique (Nelder & Mead, 1965). It is an update of a simplex via reflection, expansion, contraction and shrinkage operations and is therefore helpful when derivatives are not accessible or numerically cumbersome.

In contrast to BFGS and Nelder-Mead, GWO and DE conduct a more comprehensive exploration of the parameter space and maintain populations of candidate solutions. This may decrease the sensitivity to starting values, although it typically necessitates a greater number of likelihood evaluations. As a result, the techniques are assessed not only in terms of estimation accuracy but also in terms of computing cost and convergence dependability.

2.2.4. Implementation and comparison criteria

To reduce the impact of stochastic variation, GWO and DE are run multiple times independently for each estimate problem, and the solution with the highest log-likelihood is kept. Parameter constraints are designed to ensure that all feasible solutions fall within the allowed LTL parameter domain.

All four approaches are evaluated on the same simulated data. We assess their performance in terms of bias, mean squared error (MSE), root mean squared error (RMSE), maximised log-likelihood, convergence rate and computing time. This architecture enables the comparison of the statistical correctness and the numerical efficiency of the different optimisation approaches.

2.3. Monte Carlo Simulation Study

A Monte Carlo simulation study was conducted to evaluate the finite-sample performance of the optimization methods for maximum likelihood estimation of the Lambert–Topp–Leone regression model. The two metaheuristic methods, GWO and DE, were compared with the classical BFGS and Nelder–Mead (NM) algorithms under identical simulated samples.

2.3.1. Simulation design

Consider the regression structure

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \quad i = 1, \ldots, n, \quad (16)$$

where x_{i1} is generated from a standard normal distribution and x_{i2} from a Bernoulli distribution with probability 0.5. The linear predictor is linked to the selected positive LTL parameter through the logarithmic link defined in Section 2.

To examine different regression configurations, three parameter scenarios are considered:

$$S1 : (\beta_0, \beta_1, \beta_2) = (0.50, 0.50, -0.50), \quad (17)$$

$$S2 : (\beta_0, \beta_1, \beta_2) = (1.00, -0.75, 0.50), \quad (18)$$

$$S3 : (\beta_0, \beta_1, \beta_2) = (-0.50, 1.00, -1.00). \quad (19)$$

The remaining LTL parameter is fixed at an admissible value throughout each scenario. Sample sizes

$$n \in \{50, 100, 200, 500\} \quad (20)$$

are considered to investigate the effect of increasing information on estimation accuracy.

For each combination of sample size and parameter scenario, R = 1000 independent samples are generated from the LTL regression model. Each sample is fitted using GWO, DE, BFGS, and NM.

2.3.2. Algorithm settings

For a fair comparison, GWO and DE use the same population size and maximum number of iterations. The settings adopted in the simulation are summarized in Table 1.

Table 1. Settings of the optimization algorithms.

Method	Parameter	Setting
GWO	Population size	30
	Maximum iterations	500
DE	Population size	30
	Maximum iterations	500
	Mutation factor F	0.8
	Crossover probability CR	0.7
BFGS	Maximum iterations	500
NM	Maximum iterations	500

The same admissible parameter bounds are imposed on the population-based methods. For the stochastic algorithms, independent random initialization is employed. Convergence is declared when the change in the objective function is below a prescribed numerical tolerance or when the maximum number of iterations is reached.

2.3.3. 4.3 Performance measures

Let θ denote any true model parameter and $\hat{\theta}^{(r)}$ its estimate in replication r. The empirical bias and mean squared error are calculated as

$$\mathbf{Bias}(\hat{\theta}) = (1/R) \sum_{r=1}^{R} (\hat{\theta}^{(r)} - \theta), \qquad (21)$$

and

$$\mathbf{MSE}(\hat{\theta}) = (1/R) \sum_{r=1}^{R} (\hat{\theta}^{(r)} - \theta)^2. \qquad (22)$$

The root mean squared error is

$$\mathbf{RMSE}(\hat{\theta}) = \sqrt{\mathbf{MSE}(\hat{\theta})}. \qquad (23)$$

In addition to estimation accuracy, the algorithms are compared according to the average maximized log-likelihood, convergence rate, and mean computational time. The convergence rate is defined as

$$\mathbf{CRate} = (\text{number of successful runs} / R) \times 100\%. \quad (24)$$

Since GWO and DE are stochastic optimizers, their numerical stability is also examined across independent runs. This allows the comparison to distinguish statistical estimation accuracy from optimization reliability.

2.3.4. Evaluation strategy

With diminishing absolute bias and RMSE, an effective optimisation technique should approach the genuine parameter values as n grows. Also, when all four algorithms converge to the global maximum, the maximised log-likelihood values should be almost the same as they all optimise the same log-likelihood.

Therefore, a slightly altered parameter estimate should not be used to determine the main benefit of a metaheuristic approach. Rather, its utility is evaluated by how well it regularly finds high-quality probability answers at a reasonable computing cost, especially under difficult starting values.

3. Results and Discussion

We tested the finite-sample performance of GWO, DE, BFGS, and Nelder-Mead (NM) using R = 1000 Monte Carlo replications for each parameter scenario and sample size. The findings were evaluated for bias, RMSE, maximised log-likelihood, convergence behaviour, and computing time. The investigation found that parameter setting significantly affects the relative performance of optimisation processes.

3.1. Estimation accuracy

Table 2 shows typical RMSE findings for the shape parameter α and regression intercept $\gamma 0$. The full simulation results indicate similar trends for the other regression coefficients.

Figure 1 illustrates the RMSE behaviour of $\hat{\alpha}$ in the three simulation scenarios. The results confirm that the underlying parameter configuration substantially influences the relative efficacy of the optimisation methods. In general, the RMSE values of BFGS and Nelder–Mead are lower than those of the population-based algorithms in S1 and S2. Conversely, the relative advantage undergoes a transformation under S3, as GWO and DE, in particular, become more competitive. For instance, with n = 500, DE attains the minimal RMSE for $\hat{\alpha}$. These results suggest that no optimisation technique consistently outperforms others across all LTL parameter combinations

and underscore the possible advantages of metaheuristic search in more complex probability scenarios.

In Scenario S1, BFGS and NM usually had the lowest RMSE numbers. At n = 50, the RMSE of $\hat{\alpha}$ was 0.450 for BFGS and 0.519 for NM. For DE and GWO, it was 1.368 and 1.620, respectively. Increasing the sample size made all of the methods much better at estimating the regression values. The RMSE of $\hat{\gamma}1$ was only 0.042 to 0.051 when n = 500.

Under S2, a generally comparable pattern was noted. The matching figures for GWO and DE were 0.929 and 0.968, whereas BFGS provided an RMSE of 0.325 for α' at n = 500, followed by NM with 0.406. However, the regression slopes showed substantially smaller variations between the strategies. For instance, when n = 500, the RMSE values of $\gamma^1 1$ ranged from 0.050 to 0.057.

Table 2. RMSE of selected parameter estimates based on 1000 replications.

Scenario	n	Method	α	γ_0	γ_1	γ_2
S1	50	GWO	1.620	0.871	0.150	0.297
		DE	1.368	0.738	0.152	0.279
		BFGS	0.450	0.304	0.146	0.268
		NM	0.519	0.321	0.148	0.271
S1	500	GWO	1.648	0.900	0.043	0.129
		DE	1.503	0.810	0.051	0.104
		BFGS	0.395	0.213	0.042	0.084
		NM	0.359	0.209	0.046	0.102
S2	50	GWO	0.931	0.560	0.166	0.346
		DE	0.937	0.536	0.172	0.330
		BFGS	0.601	0.408	0.167	0.325
		NM	0.588	0.392	0.173	0.326
S2	500	GWO	0.929	0.518	0.050	0.162
		DE	0.968	0.530	0.054	0.105
		BFGS	0.325	0.180	0.050	0.096
		NM	0.406	0.219	0.057	0.117
S3	50	GWO	0.570	0.371	0.170	0.359
		DE	0.756	0.487	0.171	0.331
		BFGS	1.097	0.656	0.170	0.330
		NM	0.922	0.551	0.172	0.336
S3	500	GWO	0.469	0.266	0.052	0.150
		DE	0.349	0.205	0.055	0.113
		BFGS	0.611	0.333	0.052	0.107
		NM	0.815	0.416	0.056	0.121

Scenario S3 has noticeably different results. Metaheuristic processes outperformed other methods in estimating α and γ0 in this setting. At n = 50, GWO had the lowest RMSE for α (0.570), compared to DE (0.756), NM (0.922), and BFGS (1.097). At n = 500, DE had the lowest RMSE for α (0.349), followed by GWO (0.469), BFGS (0.611), and NM (0.815). DE also has the smallest RMSE for γ0 in this sample set.

The results show that no optimisation algorithm is superior to all parameter settings. Classical approaches are well suited to the first two situations, but population-based algorithms are more competitive when the numerical features of the likelihood surface are changed by the parameter setting.

3.2. Likelihood and numerical performance

The mean maximised log-likelihood and the percentage of runs that achieve a solution that is sufficiently similar to the best solution found among the four algorithms are presented in Table 3. The results are reported for n = 500 in order to conserve space.

The numerical reliability of the competing optimisation procedures is more clearly illustrated in Figure 2. BFGS typically demonstrates the maximum level of agreement with the best likelihood solution, particularly as the sample size increases. In contrast, GWO and DE demonstrate reduced agreement rates within the established search budget, despite the fact that their performance varies among the three parameter combinations.

Nelder-Mead has moderate efficacy although falls short of BFGS in terms of reliability across different situations. These results indicate that a thorough global exploration executed by a metaheuristic algorithm does not automatically guarantee enhanced reliability in likelihood maximisation; its effectiveness is also dependent on the population size, iteration budget, and the configuration of the LTL likelihood surface. In each of the three cases, BFGS attained the highest mean log-likelihood and exhibited the most robust alignment with the ideal solution. The agreement percentages were 97.3%, 88.7%, and 87.3% for S1, S2, and S3 with a sample size of n = 500. When the probability distribution is adequately structured and the initialisation is appropriate, this demonstrates the remarkable numerical efficiency of the traditional quasi-Newton method.

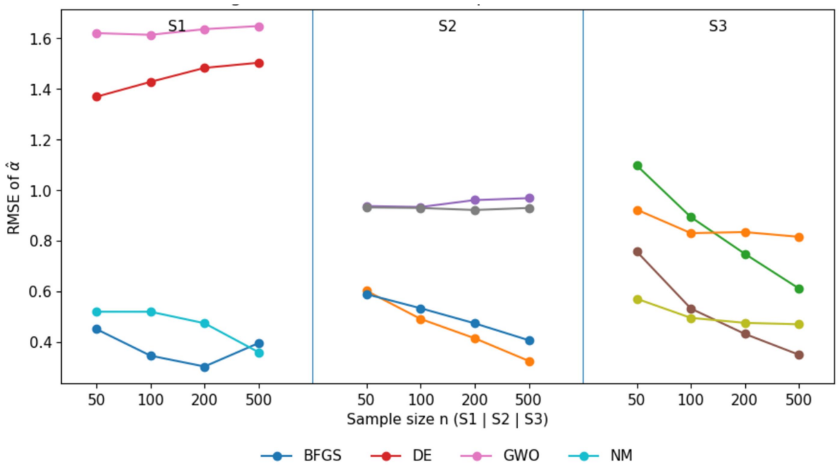


Figure 1. RMSE of the estimated LTL shape parameter α̂ for GWO, DE, BFGS, and Nelder–Mead under the three simulation scenarios and different sample sizes.

The population-based methods had a different way of computing. Even though GWO and DE didn't always get the highest likelihood under the chosen search budget, they were still able to work because they didn't need to do derivative calculations and didn't need a single fixed starting point as much. In general, DE got higher mean likelihood values than GWO, but in this implementation, GWO needed less time to run. In general, NM took the most time to run on computers of the four methods.

3.3. Overall discussion

The simulation experiment leads to three conclusions. First, estimating the regression coefficients particularly the slope parameters is much improved by increasing n. Second, especially in more challenging parameter configurations, the numerical optimiser can significantly impact the calculation of the regression intercept and LTL shape parameter. Third, neither classical optimisation nor metaheuristic optimisation are universally superior, according to the results.

Figure 3 compares the computing costs of the four optimisation strategies as sample size grows. BFGS stays computationally efficient across all sample sizes, but Nelder-Mead typically takes longer to execute. GWO outperforms DE among metaheuristic processes in the implementation used. DE's computing overhead increases significantly with sample size since its mutation, crossover, and selection operations necessitate repeated probability evaluations over a population of candidate solutions.

BFGS has the best overall likelihood performance in S1 and S2, and GWO and DE have competitive estimate accuracy in S3. Metaheuristic optimisation may be useful, depending on the circumstance. These findings show that GWO and DE are complementary global search alternatives, not substitutes. for traditional maximum likelihood optimisation. In practice of LTL regression, the comparison of the resulting likelihood values from different optimisation processes might give an additional diagnostic for the reliability of the numerical maximum likelihood solution.

Table 3. Optimization performance at n = 500.

Scenario	Method	Mean log-likelihood	Agreement (%)	Time (s)
S1	GWO	155.727	9.9	0.011
	DE	157.753	0.6	0.022
	BFGS	161.877	97.3	0.013
	NM	159.732	47.9	0.035
S2	GWO	180.295	17.7	0.014
	DE	181.062	8.3	0.029
	BFGS	181.614	88.7	0.016
	NM	179.924	31.8	0.044
S3	GWO	682.089	20.7	0.013
	DE	682.792	14.2	0.027
	BFGS	683.141	87.3	0.019
	NM	681.640	13.1	0.038

The MockJurors dataset was employed to evaluate the practical performance of the Lambert–Topp–Leone regression model. The data consist of n = 104 observations that were collected from an experimental study of mock jurors. The response variable, confidence, is scaled to the open unit interval (0, 1) and represents the juror's confidence in the selected verdict. Two explanatory factors are considered: verdict, indicating whether a two-option or three-option verdict system was employed, and conflict, indicating the presence or absence of conflicting testimonial evidence.

For the LTL regression analysis, sum coding was employed for the two categorical predictors, and the following model was fitted:

$$\log(\beta_i) = \gamma_0 + \gamma_1 V_i + \gamma_2 C_i + \gamma_3 V_i C_i, \quad (25)$$

where V_i denotes the verdict effect, C_i denotes the conflict effect, and α is treated as a common shape parameter. Thus, the complete parameter vector is

$$\psi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \alpha)^T.$$

The likelihood was optimised independently using GWO, DE, BFGS, and Nelder–Mead. Ten independent runs were conducted for the stochastic procedures, with a population size of 50 and 500 iterations. The solution with the highest probability was chosen.

3.4. Parameter estimation

Table 4. presents the estimates obtained by the four optimization strategies.

Table 4. LTL regression estimates for the MockJurors data.

Method	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\alpha}$	ℓmax
GWO	1.2995	-0.0828	-0.3315	-0.4793	2.2234	4.1452
DE	1.2995	-0.0829	-0.3316	-0.4790	2.2236	4.1452
BFGS	1.2995	-0.0829	-0.3316	-0.4790	2.2236	4.1452
NM	1.2995	-0.0829	-0.3316	-0.4790	2.2236	4.1452

Figure 4 visually substantiates the strong correlation among the four optimisation methods in the real-data application. The estimations derived from GWO, DE, BFGS, and Nelder–Mead are virtually identical for all regression coefficients and the LTL shape parameter α. All techniques provide $\hat{\alpha}$ = 2.224, comfortably situated within the permissible parameter range. Despite using quite different search strategies, this strong agreement shows that the competing algorithms converge to nearly the same maximum likelihood answer. Additionally, the outcome validates the employment of GWO and DE as separate global-search methods to confirm the numerical stability of classical likelihood optimisation in LTL regression. The MockJurors dataset was used to evaluate the empirical effectiveness of the Lambert-Topp-Leone regression model. The dataset contains 104 observations acquired from an experimental inquiry with fictitious juries. The dependent variable, confidence, is normalised to the open unit interval (0, 1) and represents the juror's certainty in the chosen decision.

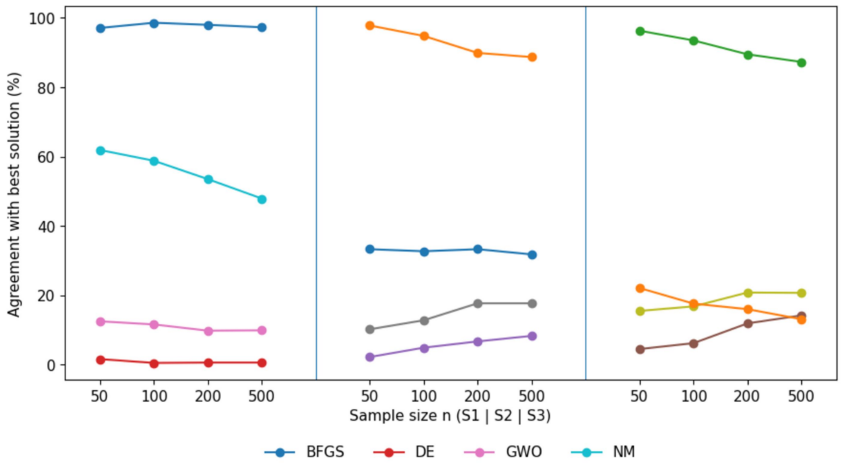


Figure 2. Optimization reliability of GWO, DE, BFGS, and Nelder–Mead across the three simulation scenarios and different sample sizes. Reliability is measured by the percentage of runs attaining a solution sufficiently close to the best likelihood value obtained among the competing optimization methods.

3.5. Optimization performance

Table 5 summarizes the numerical performance of the four optimization procedures.

Table 5. Optimization performance for the MockJurors data.

Method	Log-likelihood	AIC	BIC	Time (s)
GWO	4.14524	1.70953	14.93148	0.734
DE	4.14524	1.70952	14.93147	0.997
BFGS	4.14524	1.70952	14.93147	0.020
NM	4.14524	1.70952	14.93147	0.053

All four methods attain essentially the same maximum log-likelihood, confirming that the metaheuristic and classical procedures identify the same maximum likelihood solution for these data. Nevertheless, substantial differences occur in computational cost. BFGS is the fastest method, followed by NM, whereas the population-based GWO and DE procedures require more likelihood evaluations and consequently greater computation time.

The close agreement is nevertheless important. GWO and DE obtain the same solution without depending on local derivative information, demonstrating their ability to recover the LTL maximum likelihood solution through global population-based search. Thus, the real-data results complement the simulation experiment: metaheuristic optimization can provide an independent global-search verification of solutions obtained using classical numerical methods.

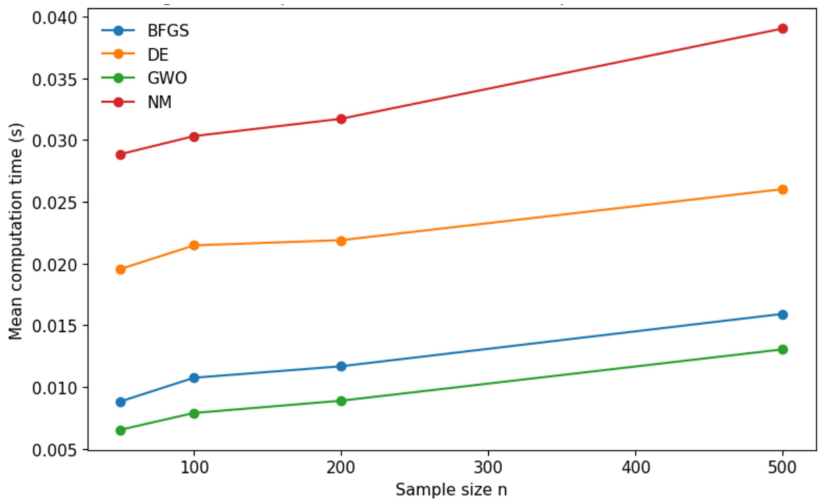


Figure 3. Mean computational time of GWO, DE, BFGS, and Nelder–Mead for different sample sizes, averaged over the three simulation scenarios.

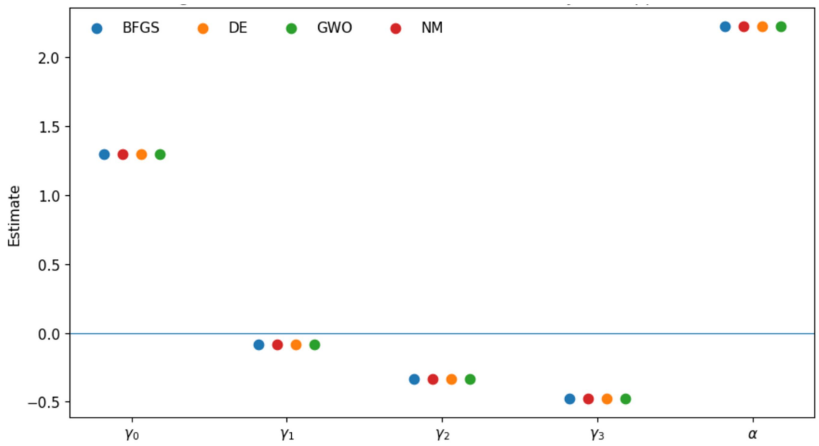


Figure 4. Maximum likelihood estimates of the LTL regression parameters for the MockJurors data obtained using GWO, DE, BFGS, and Nelder–Mead.

3.6. Comparison with beta regression

For reference, the LTL regression fit was also compared with a conventional beta regression using the same mean-model covariates. The results are shown in Table 6.

Table 6. Comparison of LTL and beta regression for the MockJurors data.

Log-likelihood	AIC	BIC
4.145	1.710	14.931
30.580	-51.161	-37.939

For this particular dataset, beta regression provides a considerably larger maximized log-likelihood and substantially smaller AIC and BIC values than the LTL regression model. Therefore, the LTL model should not be interpreted as universally superior to existing bounded-response models. Rather, its value lies in providing an alternative distributional structure whose suitability depends on the characteristics of the observed response.

From the computational perspective, however, the application provides strong evidence regarding optimization reliability: GWO, DE, BFGS, and NM converge to practically identical LTL parameter estimates. The result supports the use of metaheuristic methods as complementary global-search tools for checking the stability of maximum likelihood estimation in LTL regression.

4. Conclusions

In this paper, a Lambert-Topp-Leone regression model for bounded responses was designed and the maximum likelihood estimate was studied using two metaheuristic algorithms, Grey Wolf Optimiser (GWO) and Differential Evolution (DE), with the classical BFGS and Nelder-Mead (NM) approaches. The suggested architecture allows for the inclusion of covariate information via a log-link and maintains the necessary parameter restrictions of the LTL distribution. The Monte Carlo results showed that the choice of numerical optimizer can affect finite-sample estimation, particularly for the LTL shape parameter and regression intercept. BFGS provided the strongest overall likelihood performance in the first two simulation scenarios, whereas GWO and DE became more competitive under the third and more challenging parameter configuration. The results therefore indicate that no single optimization method is uniformly superior across all settings. Using the MockJurors data to show that GWO, DE, BFGS, and NM can all come up with pretty much the same maximum likelihood answer was another proof. The metaheuristic procedures needed more work on the computer, but they offered a separate global-search method for checking the stability of solutions found by classical optimisation. The AIC and BIC values for beta regression were higher than those for the LTL model for these data. This means that the suggested model should not be seen as a replacement for all existing bounded-response models, but rather as an addition to them. Overall, the results point to a practical method for evaluating the numerical reliability of maximum likelihood estimate in LTL regression: combining classical and metaheuristic optimisation. Alternative connection architectures, covariate-dependent form parameters, variable dispersion, and hybrid global–local optimisation techniques might all be taken into consideration in future research.

References

Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S., & Shahbaz, M. Q. (2016). Topp-Leone family of distributions: Some properties and application. *Pakistan Journal of Statistics and Operation Research*, 12(3), 443–451. https://doi.org/10.18187-/pjsor.v12i3.1458

Astorga, J. M., & Iriarte, Y. A. (2025). The Lambert-Topp-Leone distribution: An alternative for modeling proportion and lifetime data. *Frontiers in Applied Mathematics and Statistics*, 11. https://doi.org/10.3389/fams.2025.1527833

Bourguignon, M., Silva, R. B., & Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12(1), 53–68. https://doi.org/10.6339/JDS.201401_12(1).0004

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA Journal of Applied Mathematics*, 6(3), 222–231. https://doi.org/10.1093/imamat/6.3.222

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Duxbury.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2), 1–24. https://doi.org/-10.18637/jss.v034.i02

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815. https://doi.org/10.1080/0266476042000214501

Kieschnick, R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling*, 3(3), 193–213. https://doi.org/10.1191/1471082X03st053oa

Lemonte, A. J., & Cordeiro, G. M. (2011). Birnbaum–Saunders nonlinear regression models. *Computational Statistics & Data Analysis*, 55(1), 62–72. https://doi.org/10.1016/j.csda.2010.09.008

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

Nadarajah, S., & Kotz, S. (2003). Moments of some J-shaped distributions. *Journal of Applied Statistics*, 30(3), 311–317. https://doi.org/10.1080/0266476022000030084

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-40065-5

Pawitan, Y. (2001). In all likelihood: Statistical modelling and inference using likelihood. Oxford University Press. https://doi.org/10.1093/oso/9780198507659.001.0001

Price, K. V., Storn, R. M., & Lampinen, J. A. (2005). Differential evolution: A practical approach to global optimization. Springer.

Rezaei, S., Sadr, B. B., Alizadeh, M., & Nadarajah, S. (2017). Topp-Leone generated family of distributions: Properties and applications. *Communications in Statistics – Theory and Methods*, 46(6), 2893–2909. https://doi.org/10.1080-/03610926.2015.1053935

Storn, R., & Price, K. (1997). Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), 341–359. https://doi.org/10.1023/A:1008202821328

Talbi, E.-G. (2009). Metaheuristics: From design to implementation. John Wiley & Sons. https://doi.org/10.1002-/9780470496916

Topp, C. W., & Leone, F. C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association*, 50(269), 209–219. https://doi.org/10.1080/01621459.1955.10501259

Yang, X.-S. (2010). *Nature-inspired metaheuristic algorithms* (2nd ed.). Luniver Press.