

COMPARISON OF OPTIM, NLEQSLV AND MAXLIK TO ESTIMATE PARAMETERS IN SOME OF REGRESSION MODELS

Buu-Chau TRUONG¹, Van-Buol NGUYEN², Hoang-Vinh TRUONG³,
Thi Diem-Chinh HO^{2,4,*}

¹Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

²General Faculty, Binh Duong Economics & Technology University, Binh Duong City, Vietnam

³Department of Mathematics Education, Teachers College, Can Tho University, Vietnam

⁴Faculty of Mathematics and Computer Science, University of Sciences, Ho Chi Minh City, Vietnam

*Corresponding Author: Thi Diem-Chinh HO (Email: hothidiemchinh@gmail.com)

(Received: 5-Oct-2018; accepted: 16-Dec-2019; published: 31-Dec-2019)

DOI: <http://dx.doi.org/10.25073/jaec.201934.262>

Abstract. *Our main goal in this article is to present the approaches and examples of three functions in R consist of `optim`, `nleqslv` and `maxLik` function to detect the optimization solution of the estimating function in the regression models. We then compare the results with numerous sample sizes ($n=150, 300$ and 500), the execution time of R code, as well as Normal $Q - Q$ plots of three approaches through some of regression models such as the zero-inflated Binomial (ZIB) regression model, logistic regression model, the zero-inflated Poisson (ZIP) regression model and the zero-inflated Bernoulli (ZIBer) regression model. Finally, we discuss potential research directions in the coming times.*

Keywords

Optim, nleqslv, maxLik, optimization solution.

1. Introduction

The R software is developed by Ihaka and Gentleman [15] and it continues to be developed by the R Development Core Team. Software R is one of the statistical analysis tools as well as data analysis in general. In the past 10 years, R has been largely utilized by several universities around the world. This is open source software (free). It carries all the features of other existing commercial software such as SPSS, AMOS, STATA or EViews. About this regard, readers may refer in (see e.g. [3], [4], [5], [7], [11], [19], [20], [24], [26], [27], [28] and [33] etc.)

The R software contains numerous types of statistical techniques and graphics. The R software, like S, is designed around a real machine language, and it allows users to add additional features by defining new functions. S software comes before R software. S is a statistical programming language developed primarily by Chambers [6]. In this time, two books were published by the research team at Bell Laboratories: S: An Interactive Environment for Data Analysis and Graphics [1] and Extending the S System [2]. Hence there are also some vital dif-

ferences for S, but numerous codes are written in S can run without modification. Several systems in R are written in its own language, making it easy for users to follow algorithms. In order to perform computational tasks, R can connect to C, C++ and Fortran languages to be called at runtime. Proficient users can write C code to directly handle R software objects.

In several fields of statistical inference, social science and numerous other areas, the issues involve to detect the optimization solutions play an extremely crucial role in research. In statistical inference, the issue to find the optimization solution is basically in all models, for instance, time series models and regression models. In social science research, this issue typically intervenes to address numerous practical problems, for instance, transportation problems. To simplify calculation and save time in programming, several ubiquitous functions in R are proposed to detect the optimization solutions.

For the available function in R to find the optimization solutions is very diverse and abundant. For instance, The optim function is offered by Nash [21]. Eberhart and Kennedy [10] introduce the Particle Swarm Optimization (PSO) algorithm. Ter [25] provides the genetic algorithm Differential Evolution (DE) for Markov Chain Monte Carlo. The nleqslv function is developed by Hasselman [12]. The maxLik function is proposed by Henningsen and Toomet [13]. Scrucca [23] presents to a package for Genetic Algorithms (GA), etc. It has been seen that, there are many packages in R to find the optimization solutions. As our knowledge, to find the optimization solution of the estimating functions in the regression models then optim, nleqslv and maxLik function are the most used.

It can be seen that the R software is a vigorous software, open source, free download, and very simple to install. The optimization solutions of the estimating function in the regression models play an utmost important role and very profound meaning in life. Therefore it is extremely meaningful to have a study about the functions in R to find the optimization solution of the estimating function in the regression models. Although there are many authors had researched to this issue, the problem about a comparison of

the functions in R to find the optimization solution of the estimating function in the regression models has not been investigated. Our main goal in this article is to present the approaches and examples of three functions in R consist of optim, nleqslv and maxLik function. We then compare the results with numerous sample sizes, the execution time of R code, as well as Normal Q - Q plots of three approaches through some of regression models. Finally, we discuss potential research directions in the coming times.

The remainder of the paper is structured as follows. In Section 2, we review of the optimization problem as well as approaches and examples of three functions in R consist of optim, nleqslv and maxLik function. Section 3 provides the comparing about the results with several sample sizes ($n=150, 300$ and 500), the execution time of R code, as well as Normal Q - Q plots of three approaches through some of regression models. Section 4 presents advantages and disadvantages of these functions. The last section gives some concluding remarks of these functions being discussed in our paper.

2. Literature review

In this section, we first revisit of the optimization problem and we then revisit of three functions in R include optim, nleqslv and maxLik function to detect the optimization solution of the estimating function in the regression models. We now discuss about the optimization problem in the next section.

2.1. Optimization problem

Let $g(u)$ be a function of a vector u . Optimization problems are related to the task of finding u^* such that $g(u^*)$ is a local maximum (or minimum). In the case of maximization

$$u^* = \operatorname{argmax} g(u)$$

and in the case of minimization,

$$u^* = \operatorname{argmin} g(u)$$

It will be known that statistical estimation problems are optimization problems. For instance, if

g is the likelihood function and u is a vector of parameter values, then u^* is called the maximum likelihood estimator (MLE), which has several nice theoretical properties.

When g is the posterior distribution function, then u^* is a widespread bayes estimator. Other well known estimators, for example the least squares estimator in linear regression are optimums of special objective functions. Usually the available functions in R are used to find the optimization solution to minimization problems. Nevertheless, we can still utilize them to address maximization problems by multiplying (-1) to the original objective function.

We now turn on discuss to three functions in R such as `optim`, `nleqslv` and `maxLik` function. We first present to `optim` function in the next sub-section.

2.2. Optim function

The `optim` function is a widespread available function to find the optimization solution of the estimating function in the regression models. The `optim` function is first introduced by Nash [21]. This function is already included in R by default, we just need to type the function name and the right structure to use it.

The simplest structure of the `optim` function can be described as follows:

$$\text{optim}(\text{start}, g)$$

where `start` is a precursory value for the parameters to be estimated and `g` is the objective function. If we have also calculated the derivative and stored it in a function `dg`, then the syntax is

$$\text{optim}(\text{start}, g, dg, \text{method}, \text{hessian} = \text{FALSE})$$

where `dg` is a gradient of an objective function, `method`: we can select "NM" (Nelder-Mead), "CG" (Conjugate Gradients), etc, `hess` is a hessian matrix of an objective function. If `FALSE` then the result of hessian matrix disappear, if `TRUE` then we can see the result of hessian matrix in calculation. To help readers easily perform the `optim` function in practice, we provide an example about this issue.

Example 1. Assuming that one needs to employ the `optim` function to detect the maximum value of the following function:

$$g(u_1, u_2) = 50(u_1 - u_2^2)^2 + (1 - u_2)^2$$

Solution

It can be seen that

$$\frac{\partial g}{\partial u_1} = 100(u_1 - u_2^2),$$

$$\frac{\partial g}{\partial u_2} = -200u_2(u_1 - u_2^2) - 2(1 - u_2)$$

Choosing (-1,1) vector is an initial value and perform the statistical software R to write code:

```
g=function(u) {
  u1 = u[1]
  u2 = u[2]
  50 * (u1 - u2 * u2)^2 + (1 - u2)^2
}
dg = function(x) {
  u1 = u[1]
  u2 = u[2]
  c(100*(u1 - u2 * u2), -200 * u2
    * (u1 - u2 * u2) - 2 * (1 - u2))
}
optim(c(-1,1), g)
```

If we use the simplest formula of `optim` function, `optim(c(-1,1), g)`, then the result is provided as follows:

```
$par
[1] 0.9947242 0.9974764
$value
[1] 9.127087e-06
$counts
function gradient
155          NA
$convergence
[1] 0
$message
NULL
```

To see the hessian matrix, one only needs to add "hessian = TRUE" in the above formula.

```
> optim(c(-1,1), g, hessian = TRUE)
$par
[1] 0.9947242 0.9974764
```

```

$value
[1] 9.127087e-06
$count
function gradient
155      NA
$convergence
[1] 0
$message
NULL
$hessian
[,1]      [,2]
[1,] 100.0000 -199.4953
[2,] -199.4953  400.0310
    
```

In addition, we also use the derivative of the objective function in the formula of optim function. For example

```

optim(c(-1,1), g, dg, method =
      "CG", hessian = TRUE)
    
```

Furthermore, one can choose the different method in the formula of optim function

```

optim(c(-1,1), g, dg, method =
      "BFGS", hessian = TRUE)
    
```

or

```

optim(c(-1,1), g, dg, method =
      "NM", hessian = TRUE)
    
```

We now present about the nleqslv function in the next sub-section

2.3. Nleqslv function

Likewise the optim function, the nleqslv function is an ubiquitous available function in R. The nleqslv function is first developed by Hasselman [12]. The simplest structure of the nleqslv function can be illustrated as follows:

$$nleqslv(start, g)$$

where *start* is a precursory value for the parameters to be optimized and *g* is the objective function. The complete formula can be expressed as follows:

```

optim(start, g, jac = NULL, method,
      jacobian = FALSE)
    
```

where *jac* is a function to return the Jacobian for the *g* function. If not supplied numerical derivatives will be used. For *method*, we can choose *method*="Broyden" or *method*="Newton", *jacobian* is a jacobian matrix of an objective function. If FALSE then the result of jacobian matrix disappear, if TRUE then we can see the result of jacobian matrix in calculation. To help readers easily perform the nleqslv function in practice, we provide an example about this issue.

Example 2. Assuming that one needs to employ the nleqslv function to detect the maximum value of the following system of equations:

$$\begin{cases} y_1 = u_1^2 + u_2^2 - 2 \\ y_2 = exp(u_1 - 1) + u_2^3 - 2 \end{cases}$$

Solution

Choosing (2,0.5) vector is an initial value and perform the statistical software R to write code:

```

g=function(u) {
y = numeric(2)
y[1] = u[1]^2 + u[2]^2 - 2
y[2] = exp(u[1]-1) + u[2]^3 - 2
y
}
nleqslv(c(1.5,1),g,jacobian=TRUE)
    
```

If we use the simplest formula of nleqslv function, nleqslv(c(1.5,1),g), then the result is provided as follows:

```

> nleqslv(c(1.5,1),g)
$x
[1] 1.491601 1.332338
$fvec
[1] -3.346834e-11 -1.306990e-10
$termcd
[1] 1
$message
[1] "Function criterion near zero"
$scalex
[1] 1 1
$nfcnt
[1] 7
$njcnt
[1] 1
$iter
[1] 7
    
```

To see the jacobian matrix in the result, one only needs to add "jacobian = TRUE" in the above formula.

We now present about the maxLik function in the next sub-section

2.4. MaxLik function

Similarly to optim function and nleqslv function, the maxLik function is a popular available function to get the optimization solution of the estimating function in the regression models. The maxLik function is first proposed by Henningsen and Toomet [13]. The simplest structure of this function is given by:

$$\text{maxLik}(\text{logLik}, \text{start})$$

where *logLik* is the log-likelihood function of an objective function and *start* is a precursory value of parameters need to be estimated.

The complete formula can be written as follows:

$$\text{maxLik}(\text{logLik}, \text{grad} = \text{NULL}, \text{hess} = \text{NULL}, \text{start}, \text{method}, \text{constraints} = \text{NULL}, \dots)$$

where *grad* is a gradient of an objective function. If NULL, numerical gradient will be executed, *hess* is a hessian matrix of an objective function. If NULL, numerical Hessian will be performed, *method*: we can select "NM" (Nelder-Mead), "CG" (Conjugate Gradients), "BFGS" is a quasi-Newton method (Broyden-Fletcher-Goldfarb-Shanno), etc, *constraints*: if we can select NULL for unconstrained maximization.

It has been seen that: To perform a maxLik function in practice, one needs to have an objective function (a log-likelihood function) and an initial value (real value or vector). To help readers easily perform the maxLik function in practice, we provide an example about this issue. We next investigate an example in paper of Henningsen et al. [13].

Example 3. *Supposing that u is drawn from $N(2, 3)$ and the sample size is 1000. Find the parameters are estimated from this distribution.*

Solution

The log-likelihood function can be described as

follows:

$$\ln(L(u; \mu, \sigma)) = -\frac{N \ln(2\pi)}{2} - N \ln(\sigma) - \sum_{i=1}^n \frac{(u_i - \mu)^2}{2\sigma^2}$$

Choosing (2,3) vector is an initial value and perform the statistical software R to write code:

```
>u = rnorm(1000, mean = 2, sd = 3)
>loglikelihood.fun = function(par){
mu = par[1]
sigma = par[2]
sum(dnorm(u, mean = mu, sd = sigma,
log = TRUE))}
>MLE= maxLik(logLik =
loglikelihood.fun, start =
c(mu = 2, sigma = 3))
```

To thoroughly see the adequate results of MLE, one can employ the following command: summary(MLE). For simplicity's sake, let coef be parameters need to be estimated and stdEr be standard errors, one can utilize coef(MLE) and stdEr(MLE) to look at the result of the parameters need to be estimated and its standard error.

```
> summary(MLE)
> coef(MLE)
mu          sigma
1.872786    3.043683
> stdEr(MLE)
mu          sigma
0.09652925  0.06796909
```

3. Comparing the performance of three approaches

It will be known that, the regression models play an extremely important and significant role in statistics, economics, finance, etc. If the data sets are used to apply to the regression model contains too many zeroes, traditional regression models such as logistic, Poisson, Binomial, etc. will no longer be effective in research because the result in estimates will be biased. Thus the zero-inflated models have been proposed to overcome

this problem. Corresponding to the traditional regression models, we just need to add the word "zero-inflated", then we will have the new regression models. For instance, the zero-inflated Bernoulli (ZIBer) is proposed to improve the logistics regression model when the outcome variable contains too many zeroes. The zero-inflated Poisson (ZIP) is proposed to improve the Poisson regression model when the outcome variable contains too many zeroes. The zero-inflated Binomial (ZIB) is proposed to improve the Binomial regression model when the outcome variable contains too many zeroes, etc. We investigate the problem of estimating parameter of the ZIB model in the next example.

Example 4. *In this example, we compare the results with numerous sample sizes ($n=150, 300$ and 500), the execution time of R code, as well as Normal Q - Q plots of three approaches through the ZIB model. This model is introduced in Diallo et al. [8].*

Generating the data set as follows:

$$\text{logit}(\pi_i) = \beta_1 W_{i1} + \beta_2 W_{i2} + \beta_3 W_{i3} + \beta_4 W_{i4} + \beta_5 W_{i5} + \beta_6 W_{i6} + \beta_7 W_{i7} ,$$

and

$$\text{logit}(p_i) = \theta_1 X_{i1} + \theta_2 X_{i2} + \theta_3 X_{i3} + \theta_4 X_{i4} + \theta_5 X_{i5} ,$$

where $W_{i1} = 1$, $W_{i2} \sim \mathcal{N}(0, 1)$, $W_{i3} \sim \mathcal{U}(2, 5)$, $W_{i4} \sim \mathcal{N}(1, 1.5)$, $W_{i5} \sim \mathcal{E}(1)$, $W_{i6} \sim \mathcal{B}(1, 0.3)$ and $W_{i7} \sim \mathcal{N}(-1, 1)$ are independently. Let $X_{i1} = 1$, $X_{i4} \sim \mathcal{N}(-1, 1)$ and $X_{i5} \sim \mathcal{B}(1, 0.5)$ are independently. In this model, assuming that $X_{i2} = W_{i2}$ and $X_{i3} = W_{i6}$ and choosing starting values are: $\beta = (-0.3, 1.2, 0.5, -0.75, -1, 0.8, 0)^T$ and $\theta = (-0.55, -0.7, -1, 0.45, 0)^T$.

We investigate numerous sample sizes ($n = 150, 300, 500$) and $h_i \in \{4, 5, 6\}$. The numbers h_i are allowed to change across subjects.

Let

$$(k_4, k_5, k_6) = (\text{card}\{i : h_i = 4\}, \text{card}\{i : h_i = 5\}, \text{card}\{i : h_i = 6\})$$

When $n = 150$, choosing $(k_4, k_5, k_6) = (60, 50, 40)$. When $n = 300$, selecting $(k_4, k_5, k_6) = (120, 100, 80)$ and with $n = 500$, using $(k_4, k_5, k_6) = (200, 170, 130)$.

Performing above values, the average proportion of zero-inflation in our data set is 25%. The number of repetitions in simulation is chosen $N = 1000$ times and find the maximum likelihood estimation (MLE) $\hat{\alpha}_n = (\hat{\beta}_n^T, \hat{\theta}_n^T)^T$.

In this study, we perform three methods: optim function, nleqslv function and maxLik function to estimate parameters. These results are presented in Tab. 1, 2 and 3, respectively.

Note: **bias** is the distance between the estimator and its true value, **SD** is the standard deviation, **ASE** is the asymptotic standard error, and **CP** is the coverage probability of 95% level confidence intervals. All results are simulated with the number of repetition is $N = 1000$.

Tab. 1: The results by performing the optim function.

The results of estimator of β								
n		$\hat{\beta}_n$						
		$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\beta}_{7,n}$
150	bias	-0.0123	0.0279	0.0131	-0.0242	-0.0319	0.0277	0.0022
	SD	0.5681	0.1680	0.1419	0.1023	0.1641	0.2739	0.1259
	ASE	0.5560	0.1638	0.1451	0.0988	0.1589	0.2721	0.1237
	CP	0.9461	0.9421	0.9432	0.9450	0.9471	0.9477	0.9473
	$l(CI)$	2.1661	0.6358	0.5692	0.3881	0.6232	1.0583	0.4854
	300	bias	-0.0131	0.0156	0.0077	-0.0112	-0.0154	0.0079
SD		0.3879	0.1143	0.1012	0.0723	0.1134	0.1882	0.0865
ASE		0.3787	0.1131	0.0988	0.0692	0.1089	0.1864	0.0865
CP		0.9484	0.9489	0.9494	0.9515	0.9432	0.9465	0.9433
$l(CI)$		1.4864	0.4434	0.3922	0.2678	0.4347	0.7179	0.3318
500		bias	0.0024	0.0126	0.0042	-0.0078	-0.0089	0.0076
	SD	0.2931	0.0864	0.0782	0.0524	0.0831	0.1434	0.0675
	ASE	0.2931	0.0849	0.0775	0.0532	0.0843	0.1424	0.0667
	CP	0.9523	0.9456	0.9478	0.9487	0.9489	0.9442	0.9532
	$l(CI)$	1.1389	0.3367	0.2988	0.2056	0.3189	0.5487	0.2559

The results of estimator of θ						
n		$\hat{\theta}_n$				
		$\hat{\theta}_{1,n}$	$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$	$\hat{\theta}_{4,n}$	$\hat{\theta}_{5,n}$
150	bias	-0.0763	-0.0512	-0.0743	0.0522	0.0083
	SD	0.5215	0.3586	0.7379	0.3178	0.5817
	ASE	0.5043	0.3476	0.7754	0.3086	0.5789
	CP	0.9615	0.9543	0.9656	0.9582	0.9577
	$l(CI)$	1.9376	1.3265	2.8954	1.1763	2.2257
	300	bias	-0.0374	-0.0216	-0.0743	0.0267
SD		0.3412	0.2279	0.5043	0.2045	0.3819
ASE		0.3376	0.2253	0.4965	0.2022	0.3731
CP		0.9516	0.9523	0.9588	0.9517	0.9589
$l(CI)$		1.2824	0.8786	1.8917	0.7863	1.4884
500		bias	-0.0188	-0.0074	-0.0388	0.0164
	SD	0.2489	0.1743	0.3692	0.1563	0.2954
	ASE	0.2488	0.1734	0.3652	0.1553	0.2877
	CP	0.9512	0.9467	0.9515	0.9534	0.9545
	$l(CI)$	0.9796	0.6664	1.4286	0.5894	1.1234

Tab. 2: The results by performing the nleqslv function.

The results of estimator of β								
n		$\hat{\beta}_n$						
		$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\beta}_{7,n}$
150	bias	-0.0145	0.0299	0.0146	-0.0256	-0.0372	0.0295	0.0022
	SD	0.5685	0.1649	0.1443	0.1028	0.1674	0.2767	0.1296
	ASE	0.5539	0.1659	0.1496	0.0978	0.1566	0.2734	0.1267
	CP	0.9474	0.9423	0.9465	0.9487	0.9477	0.9473	0.9501
	$l(CI)$	2.1589	0.6378	0.5691	0.3795	0.6267	1.0594	0.4856
	300	bias	-0.0178	0.0167	0.0077	-0.0173	-0.0156	0.0093
SD		0.3879	0.1135	0.1123	0.0747	0.1207	0.1889	0.0864
ASE		0.3788	0.1145	0.0958	0.0694	0.1067	0.1894	0.0867
CP		0.9484	0.9488	0.9495	0.9517	0.9436	0.9474	0.9445
$l(CI)$		1.4869	0.4438	0.3946	0.2675	0.4348	0.7149	0.3336
500		bias	0.0106	0.0156	0.0079	-0.0097	-0.0084	0.0037
	SD	0.2931	0.0863	0.0784	0.0523	0.08247	0.1439	0.0696
	ASE	0.2945	0.0874	0.0798	0.0533	0.0857	0.1447	0.0684
	CP	0.9544	0.9496	0.9469	0.9468	0.9447	0.9485	0.9547
	$l(CI)$	1.1385	0.3385	0.2987	0.2059	0.3186	0.5483	0.2559

The results of estimator of θ						
n		$\hat{\theta}_n$				
		$\hat{\theta}_{1,n}$	$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$	$\hat{\theta}_{4,n}$	$\hat{\theta}_{5,n}$
150	bias	-0.0768	-0.0519	-0.0767	0.0579	0.0088
	SD	0.5216	0.3588	0.7379	0.3189	0.5823
	ASE	0.5056	0.3479	0.7756	0.3083	0.5785
	CP	0.9619	0.9545	0.9662	0.9591	0.9567
	$l(CI)$	1.9394	1.3267	2.8939	1.1785	2.2274
	300	bias	-0.0378	-0.0215	-0.0756	0.0279
SD		0.3436	0.2279	0.5056	0.2124	0.3794
ASE		0.3468	0.2326	0.4874	0.2127	0.3974
CP		0.9497	0.9558	0.9588	0.9516	0.9584
$l(CI)$		1.2848	0.8785	1.8914	0.7848	1.4849
500		bias	-0.0187	-0.0078	-0.0394	0.0164
	SD	0.2485	0.1763	0.3695	0.1566	0.2939
	ASE	0.2453	0.1768	0.3629	0.1583	0.2849
	CP	0.9533	0.9491	0.9512	0.9506	0.9507
	$l(CI)$	0.9789	0.6663	1.4294	0.5888	1.1248

Tab. 3: The results by performing the maxLik function.

The results of estimator of β								
n		$\hat{\beta}_n$						
		$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\beta}_{7,n}$
150	bias	-0.0039	0.0179	0.0149	-0.0229	-0.0373	0.0249	0.0124
	SD	0.5708	0.1695	0.1469	0.1026	0.1648	0.2753	0.1285
	ASE	0.5563	0.1697	0.1458	0.0988	0.1539	0.2806	0.1269
	CP	0.9489	0.9426	0.9468	0.9454	0.9474	0.9469	0.9474
	$l(CI)$	2.1673	0.6343	0.5675	0.3886	0.6263	1.0585	0.4876
	300	bias	-0.0049	0.0173	0.0153	-0.0186	-0.0194	0.0128
SD		0.3869	0.1185	0.1049	0.0714	0.1154	0.1875	0.0897
ASE		0.3788	0.1164	0.0986	0.0637	0.1025	0.1814	0.0884
CP		0.9464	0.9485	0.9436	0.9540	0.9406	0.9497	0.9437
$l(CI)$		1.4887	0.4425	0.3958	0.2646	0.4349	0.7195	0.3348
500		bias	0.0174	0.0219	0.0160	-0.0097	-0.0074	0.0097
	SD	0.2958	0.0883	0.0794	0.0538	0.0895	0.1449	0.0697
	ASE	0.2939	0.0875	0.0783	0.0569	0.0886	0.1448	0.0675
	CP	0.9508	0.9488	0.9458	0.9475	0.9474	0.9497	0.9565
	$l(CI)$	1.1346	0.3384	0.2986	0.2083	0.3138	0.5464	0.2583

The results of estimator of θ						
n		$\hat{\theta}_n$				
		$\hat{\theta}_{1,n}$	$\hat{\theta}_{2,n}$	$\hat{\theta}_{3,n}$	$\hat{\theta}_{4,n}$	$\hat{\theta}_{5,n}$
150	bias	-0.0748	-0.0534	-0.0764	0.0574	0.0184
	SD	0.5226	0.3584	0.7379	0.3197	0.5823
	ASE	0.5047	0.3495	0.7763	0.3096	0.5784
	CP	0.9507	0.9538	0.9606	0.9518	0.9539
	$l(CI)$	1.9374	1.3298	2.8995	1.1783	2.2258
	300	bias	-0.0275	-0.0153	-0.0097	0.0275
SD		0.3458	0.2296	0.5064	0.2085	0.3811
ASE		0.3375	0.2265	0.4993	0.2055	0.3787
CP		0.9514	0.9507	0.9537	0.9526	0.9505
$l(CI)$		1.2847	0.8787	1.8926	0.7857	1.4894
500		bias	-0.0096	-0.0131	-0.0265	0.0273
	SD	0.2488	0.1757	0.3696	0.1576	0.2974
	ASE	0.2489	0.1726	0.3674	0.1596	0.2885
	CP	0.9512	0.9474	0.9515	0.9509	0.9538
	$l(CI)$	0.9785	0.6686	1.4239	0.5885	1.1263

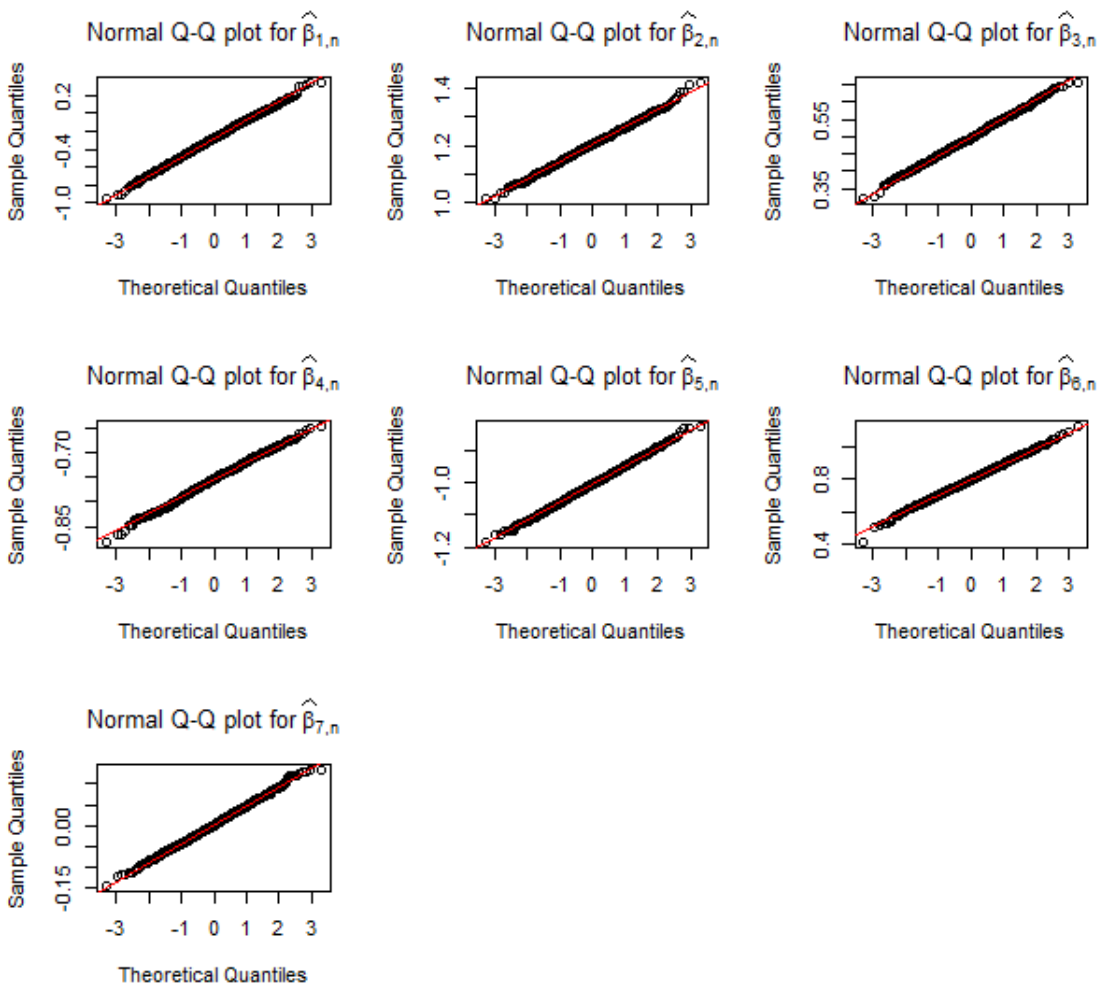


Fig. 1: Normal Q - Q plots for $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{7,n}$ with results are gained from the optim function ($n = 500$).

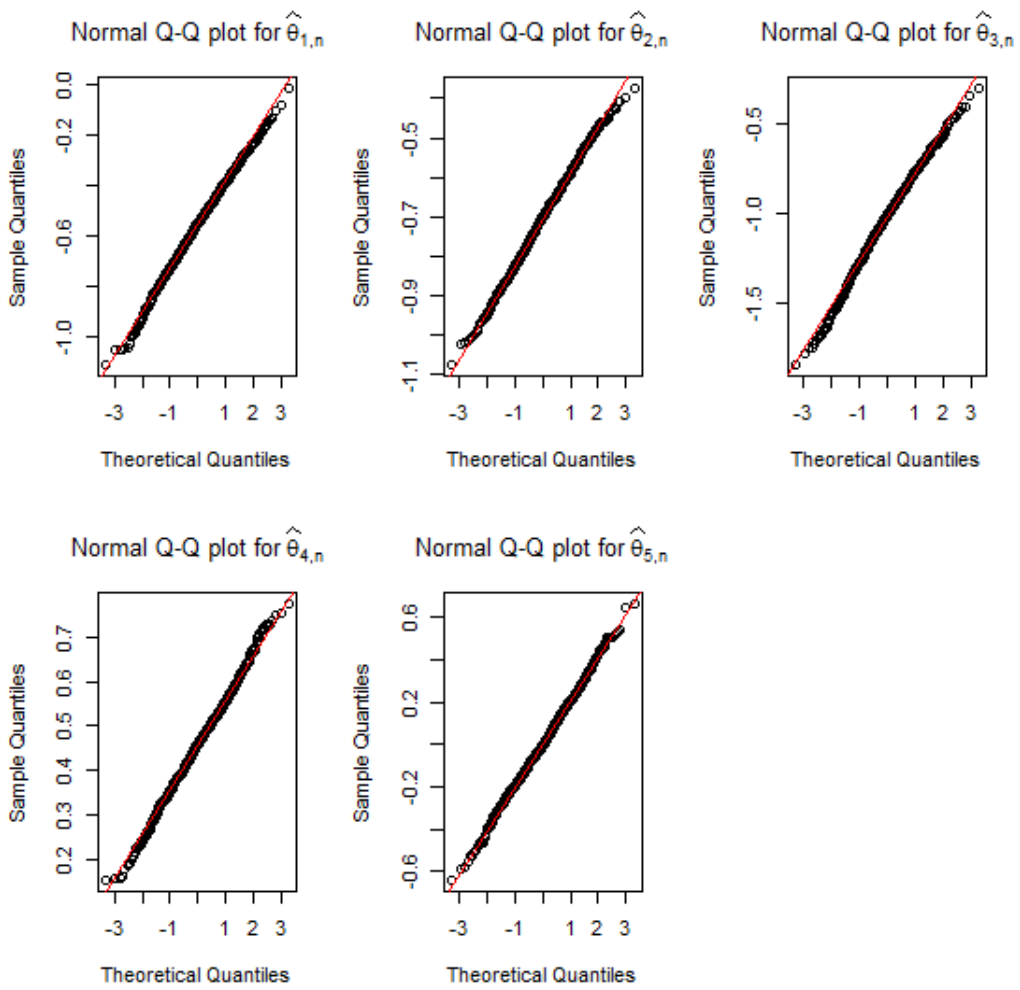


Fig. 2: Normal Q - Q plots for $\hat{\theta}_{1,n}, \dots, \hat{\theta}_{5,n}$ with results are attained by the optim function ($n = 500$).

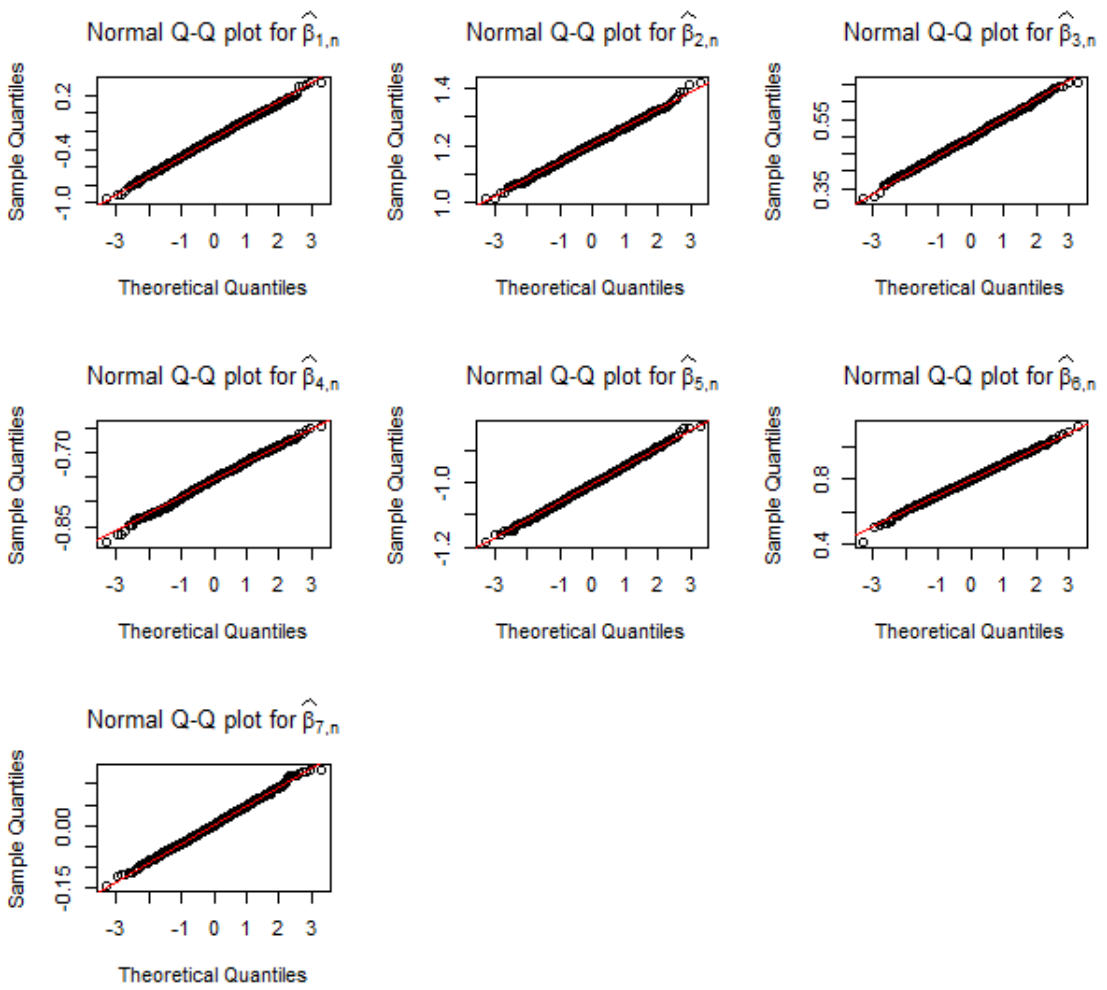


Fig. 3: Normal Q - Q plots for $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{7,n}$ with results are gained from the nleqslv function ($n = 500$).

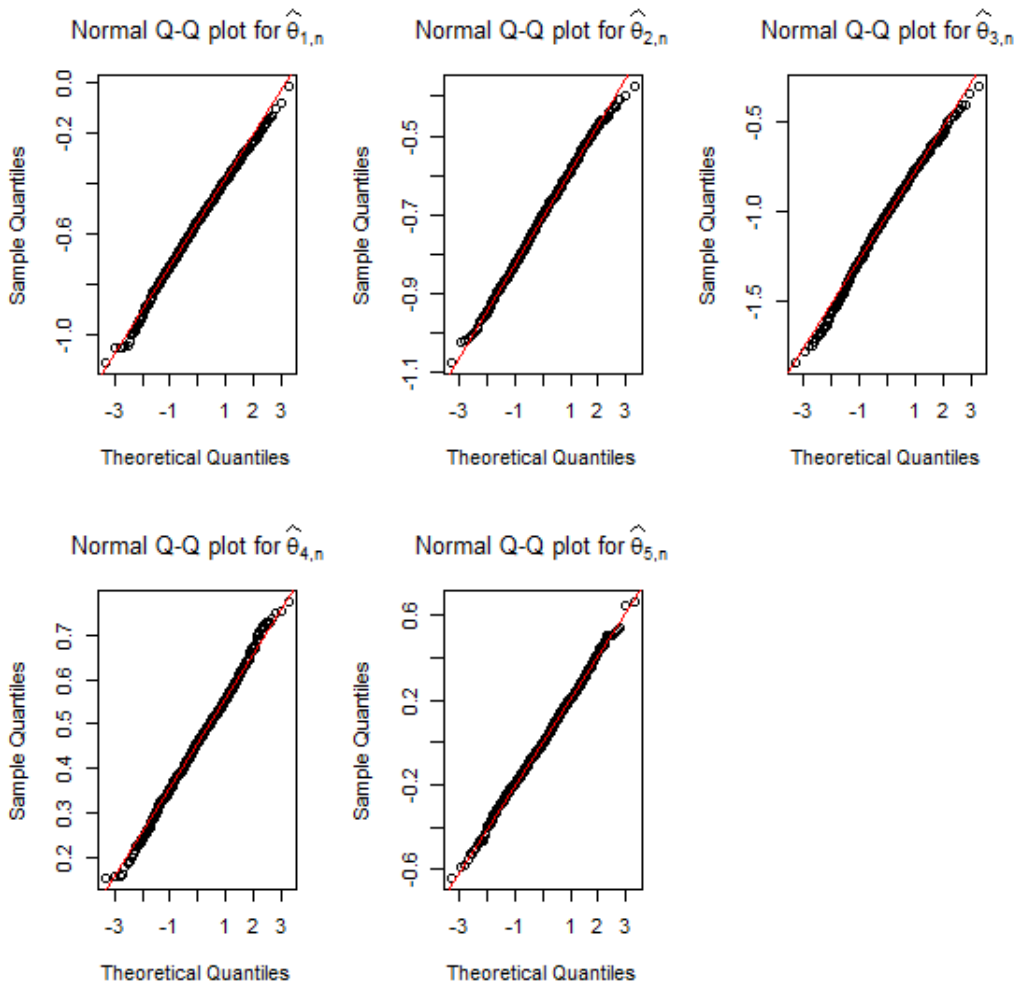


Fig. 4: Normal Q - Q plots for $\hat{\theta}_{1,n}, \dots, \hat{\theta}_{5,n}$ with results are attained by the nleqslv function ($n = 500$).

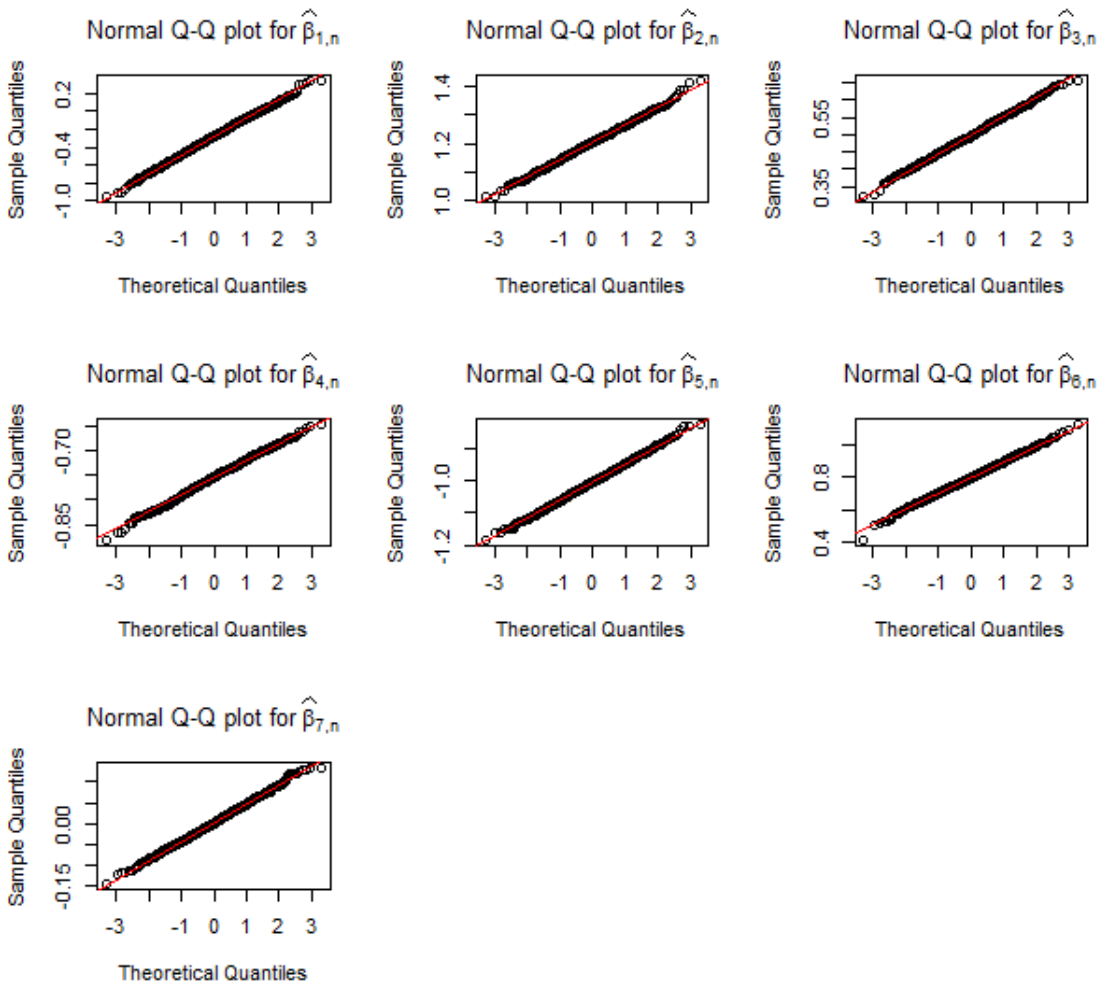


Fig. 5: Normal Q - Q plots for $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{7,n}$ with results are gained from the maxLik function ($n = 500$).

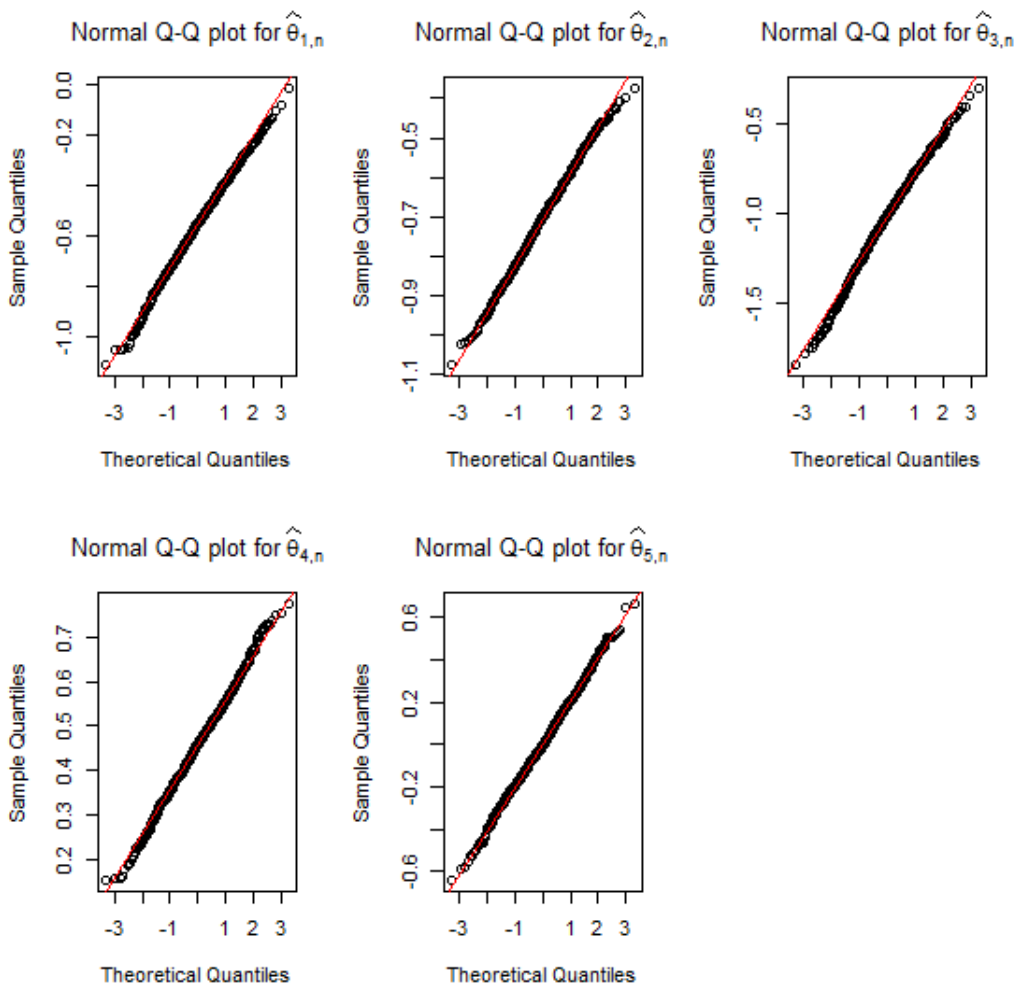


Fig. 6: Normal Q - Q plots for $\hat{\theta}_{1,n}, \dots, \hat{\theta}_{5,n}$ with results are attained by the maxLik function ($n = 500$).

It can be observed from Tab. 1, 2 and 3 that, the biases of estimators are very small, the values of SD and ASE are very close and the values of CP are very close to 0.95. These prove that our estimated results are very dependable. That is a consistent estimator. Moreover, it can be seen that the bias, SE, SD, and $l(CI)$ of all estimators decrease as the sample size increases.

About the results, the authors in paper of Di-allo et al. [8] used a maxLik function to study simulation. The results in this paper is executed by using three approaches: optim, nleqslv and maxLik function are offered in Tab. 1, 2 and 3, respectively. It can be observed from the above results of three approaches most are the same. Besides, it can be seen that the normal Q - Q plots are provided in Fig. 1-6 respectively that the Gaussian approximation of the distribution of the MLE in the ZIB model is reasonably satisfied.

To compare the execution time of R code, it is required that the data set is considered in the study needs to be the same. We utilized the Dell desktop computer is configured with Intel Core i5, 8GB of RAM, 1TB of hard drive to check the time to run code to output the result of there methods. To get the above results, for $n=150$, it takes 15 minutes for optim function, 20 minutes for nleqslv function and 30 minutes for maxLik function. For $n=300$, it takes 28 minutes for optim function, 35 minutes for nleqslv function and 55 minutes for maxLik function. For $n=500$, it takes 50 minutes for optim function, 60 minutes for nleqslv function and 70 minutes for maxLik function. Thus it can be said that the execution time of R code of the methods being discussed in the paper then the optim function is the fastest.

In example 4, we investigate the data set of Y is count data. In the next example, we aim to consider the data set of Y is binary data. Thus we present the problem of estimating parameter of the logistic model in the next example.

Example 5. *Considering the logistic regression model as follows:*

$$P(Y_i = 1) = \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} = H(\beta_0 + \beta_1 X_i)$$

where

- Y is a binary variable, $Y \in \{0, 1\}$
- X is a covariates
- β_0 and β_1 are parameters need to be estimated

The likelihood function of logistic regression model has the form

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n (P(Y_i = 1))^{Y_i} (1 - P(Y_i = 1))^{1-Y_i} \\ &= \prod_{i=1}^n \left(\frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{Y_i} \\ &\quad \times \left(1 - \frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{1-Y_i} \\ &= \prod_{i=1}^n \left(\frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{Y_i} \\ &\quad \times \left(\frac{1}{1 + e^{\beta_0 + \beta_1 X_i}} \right)^{1-Y_i} \end{aligned}$$

The log-likelihood function of logistic regression model has the form

$$\begin{aligned} \ell(\beta) &= \ln [L(\beta)] \\ &= \sum_{i=1}^n \left[Y_i \ln \left(\frac{e^{\beta_0 + \beta_1 X_i}}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \right] \\ &\quad + \sum_{i=1}^n \left[(1 - Y_i) \ln \left(\frac{1}{1 + e^{\beta_0 + \beta_1 X_i}} \right) \right] \\ &= \sum_{i=1}^n [Y_i \ln (e^{\beta_0 + \beta_1 X_i}) - \ln (1 + e^{\beta_0 + \beta_1 X_i})] \\ &= \sum_{i=1}^n [Y_i (\beta_0 + \beta_1 X_i) - \ln (1 + e^{\beta_0 + \beta_1 X_i})] \end{aligned}$$

For this example, we also investigate the results with several sample sizes ($n=150, 300$ and 500), the execution time of R code, as well as Normal Q - Q plots of three approaches. Our results show that the results of the three approaches are almost the same and the run time of the R code to get the result of the optim function is the fastest. To save space, we do not provide results in here.

For comparison purposes, we also investigate some other regression models such as the zero

inflated Poisson (ZIP) model and the zero inflated Bernoulli (ZIBer) model. The detailed formula of the likelihood function, log-likelihood function of these models may refer in Lukusa et al. [17] and Diop et al. [9], respectively. To save space, we do not provide results for this section. Our results obtained indicate that the results of the three approaches are almost the same and the run time of the R code to get the result of the optim function is the fastest.

We discuss the advantages and disadvantages of optim, nleqslv and maxLik functions in the next section.

4. Advantages and Disadvantages of these functions

Three functions in R such as optim, nleqslv and maxLik functions are the most used to find optimization solutions of the estimating function in the regression models. These functions can be employed to find precise values of the parameters need to be estimated when the data set have no missing values or errors in measurement. This is the main advantage of these functions. Moreover, these functions are easy to use and involve simple structures. Nevertheless, in practice, it is usually the case that data sets are faced up with missing or incomplete values, which is the principal disadvantage of these functions.

Cases of incomplete or missing data are an ubiquitous problem that is frequently encountered in, for instance, in the economics, technology and engineering fields, etc. This issue occurs for numerous reasons, for example, respondents not answering certain items in survey questions, non-acquiesce and inexplicable answers, etc. (for further details, see Schafer and Graham [22]).

It is clear that it is both significant and inevitableness to extend three functions in R such as optim, nleqslv and maxLik functions to be useful when the data set comprises missing values or errors in measurement. Pho and Nguyen [29] reviewed of three widespread approaches to handle the problems with missing data. Pho et

al. [31] also mentioned some methods of solving problems that contain missing data. For further details involving this issue, reference is made to Horton and Kleinman [14], Little [16], Mahmoudi et al. [18], Pho et al. [30] and [32], etc.

About the optim, nleqslv and maxLik functions. There are many optimization methods that can be chosen in these functions. In this paper, we also try to do the other optimization methods. However, our results need to satisfies three conditions: bias < 5%, the value of SD and ASE is very close and CP is very near 0.95 to our estimator is a consistent estimator. Thus if we choose the other optimization methods, then may be our estimator is not a consistent estimator. Hence we choose the suitable method in functions (optim, nleqslv and maxLik) to have a consistent estimator.

5. Conclusions

In this article, we have presented the approaches and examples of the three functions in R such as optim, nleqslv and maxLik functions. We then compare the results with many sample sizes ($n=150, 300$ and 500), the execution time of R code, as well as Normal Q - Q plots of three approaches through some of regression models such as the zero-inflated Binomial (ZIB) regression model, logistic regression model, the zero-inflated Poisson (ZIP) regression model and the zero-inflated Bernoulli (ZIBer) regression model. Our results obtained indicate that the results of the three approaches are almost the same and the execution time of R code of the optim function is the fastest.

References

- [1] Becker, R. A., & Chambers, J. M. (1984). S: an interactive environment for data analysis and graphics. CRC Press.
- [2] Becker, R. A., & Chambers, J. M. (1985). Extending the S system. Chapman and Hall/CRC.

- [3] Boehm, B., Rombach, H. D., & Zelkowitz, M. V. (Eds.). (2005). *Foundations of empirical software engineering: the legacy of Victor R. Basili*. Springer Science & Business Media.
- [4] Calenge, C. (2006). The package “adehabitat” for the R software: a tool for the analysis of space and habitat use by animals. *Ecological modelling*, 197(3-4), 516-519.
- [5] Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372-404.
- [6] Chambers, J. M. (1981). Some thoughts on expert software. In *Computer science and statistics: proceedings of the 13th symposium on the interface* (pp. 36-40). Springer, New York, NY.
- [7] Cruz, C. D. (2016). Genes Software-extended and integrated with the R, Matlab and Selegen. *Acta Scientiarum. Agronomy*, 38(4), 547-552.
- [8] Diallo, A. O., Diop, A., & Dupuy, J. F. (2017). Asymptotic properties of the maximum-likelihood estimator in zero-inflated binomial regression. *Communications in Statistics-Theory and Methods*, 46(20), 9930-9948.
- [9] Diop, A., Diop, A., & Dupuy, J. F. (2011). Maximum likelihood estimation in the logistic regression model with a cure fraction. *Electronic Journal of Statistics*, 5, 460-483.
- [10] Eberhart, R., & Kennedy, J. (1995). Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks*, 4, 1942-1948.
- [11] Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage Publications.
- [12] Hasselman, B. (2009). nleqslv: Solve systems of non linear equations. R package version, 1.
- [13] Henningsen, A., & Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. *Computational Statistics*, 26(3), 443-458.
- [14] Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61(1), 79-90.
- [15] Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3), 299-314.
- [16] Little, R. J. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association*, 87(420), 1227-1237.
- [17] Lukusa, T. M., Lee, S. M., & Li, C. S. (2016). Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. *Metrika*, 79(4), 457-483.
- [18] Mahmoudi, M. R., Heydari, M. H., Avazzadeh, Z., & Pho, K. H. (2020). Goodness of fit test for almost cyclostationary processes. *Digital Signal Processing*, 96, 102597.
- [19] Menezes, I., Zwiegelhaar, J., Moraes, E., & Mendy, J. (2019). Applying network analysis to measure organizational behaviors using R software.
- [20] Murrell, P. (2018). *R graphics*. CRC Press.
- [21] Nash, J. C. (1979). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*. Adam Hilger, Bristol. Chap, 3, 30-48.
- [22] Schafer, J.L., & Graham, J.W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, 7(2), 147.
- [23] Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37.
- [24] Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software*.

- [25] Ter Braak, C. J. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3), 239-249.
- [26] Tian, Y., & Pho, K. H. (2019). A statistical view to study the aphorisms in Nahj al-Balaghah. *Digital Scholarship in the Humanities*.
- [27] Truong, B. C., Pho, K. H., Nguyen, V. B., Tuan, B. A., & Wong, W. K. (2019). Graph Theory and Environmental Algorithmic Solutions to Assign Vehicles: Application to Garbage Collection in Vietnam. *Advances in Decision Sciences*, 23(2), 1-35.
- [28] Tuan, B. A., Pudikova, G. N., Mahmoudi, M. R., & Pho, K. H. (2019). Statistical approaches in literature: Comparing and clustering the alternatives of love in Divan of Hafiz. *Digital Scholarship in the Humanities*.
- [29] Pho, K.H., & Nguyen, V.T. (2018). Comparison of Newton-Raphson algorithm and Maxlik function. *Journal of Advanced Engineering and Computation*, 2(4), 281-292.
- [30] Pho, K.H., Ho, T.D.C., Tran, T.K., & Wong, W.K. (2019). Moment generating function, expectation and variance of ubiquitous distributions with applications in decision sciences: A review. *Advances in Decision Sciences*, 23(2), 1-85.
- [31] Pho, K.H., Ly, S., Ly, S., & Lukusa, T.M. (2019). Comparison among Akaike information criterion, Bayesian information criterion and Vuong's test in model selection: A case study of violated speed regulation in Taiwan. *Journal of Advanced Engineering and Computation*, 3(1), 293-303.
- [32] Pho, K.H., Tran, T.K., Ho, T.D.C., & Wong, W.K. (2019). Optimal solution techniques in decision sciences: A review. *Advances in Decision Sciences*, 23(1), 1-47.
- [33] Verzani, J. (2018). *Using R for introductory statistics*. Chapman and Hall/CRC.

About Authors

Buu-Chau TRUONG is a lecturer at Faculty of Mathematics and Statistics in Ton Duc Thang University, Ho Chi Minh City, Vietnam. Her currently research interests include Time Series Analysis, Bayesian Inference, Data Analysis, Applied Statistics, Financial Risk Management and Financial Mathematics.

Van-Buol NGUYEN is a lecturer at General Faculty, Binh Duong Economics & Technology University, Binh Duong City, Vietnam His research interests are Data mining, Optimization and Numerical Analysis.

Hoang-Vinh TRUONG is Master student in Department of Mathematics Education, Teachers College, Can Tho University, Vietnam. His research interests are Data mining, Optimization and Numerical Analysis.

Thi Diem-Chinh HO is Master student in Faculty of Mathematics and Computer Science, University of Sciences, Ho Chi Minh City, Vietnam and she is a lecturer in General Faculty, Binh Duong Economics & Technology University, Binh Duong City, Vietnam. Her research interests are Applied Statistics, Optimization and Financial Mathematics.