ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 10, No. 4, July 2022, pp.112-119

Bayesian estimation for the Tukey GH distribution with an application

M. F. Al-Saadony [1], Bahr Kadhim Mohammed [1], Hameedah Naeem Melik [1]
[1] University of Al-Qadisiyah, Iraq

ABSTRACT

Tukey GH is a transformed normal distribution, having two parameters (g,h). The g parameter represents skew measuring, and the h represents kurtosis measuring. Our motivation is to affect these parameters on the behavior for a simulation data and real data, such as the Iraqi stock market Index (ISX60) and the Standard and Poor's (SP500) index using the Bayesian framework. Then, our aim is to find the estimation for the parameters in this distribution using empirical and parametric Bayesian methods, and study the effective on the simulation and real data. The simulation study will be shown the behavior of the parameters with a different number of sampling and prior distributions. We will use the real data from ISX60 and SP500 index.

Keywords: Tukey GH distribution, skewness, Kurtosis, Empirical Bayesian method, Parametric Bayesian Method

Corresponding Author:

M. F. Al-Saadony
University of Al-Qadisiyah
Al-Diwaniyah, Iraq
E-mail: muhannad.alsaadony@qu.edu.iq

1. Introduction

The Tukey gh distribution is transformation of standard normal distribution, when the function is nonlinear. The tukey gh distribution deals with skewness and kurtosis. The normal distribution may be having some problems such as heavy tails and kurtosis. Therefore, the tukey gh distribution have taken four parameters, the first and second parameters (a,b), the third parameter g which represents the skewness aspect and the fourth parameter h which represents the kurtosis aspect [1, 2, 3].

Our motivation is to attempt to do some statistical Inference for these parameters. It is to study the effective on the behavior of the tukey gh distribution. In this paper, we will estimate g and h parameters using some different Bayesian methods such as Empirical and Parametric Bayesian methods. In literature review, Tukey (1977) discussed TGH distribution with some properties of it [1]. There are many articles about the TGH distribution and its properties but we will focus on the articles which is represented the estimation the parameters of the TGH distribution within the previous. Xu and Genton (2015) suggested the efficient maximum likelihood inference for TGH distribution [4]. Bee and Trapin (2016) presented the novel approach of Approximate Bayesian Computation (ABC) for the gh distribution using simulation and real data [5]. Linda Mostel et.al.(2019) estimated the parameters of the tukey-type distribution using Non- Bayeian methods as comparative analysis [6].

The paper is organized as follows: Section 2 reviews the Tukey gh distribution and its properties. Section 3 introduces the Empirical Bayesian method. Section 4 present Parametric Bayesian method to estimate our parameters. Section 5 and 6, respectively, give the results of our simulation and real data study. Finally, Section 7 we present some issues for further work.

2. Tukey g-h distribution

Tukey (1977) presented a new generalized of g-h distribution to study the effect the skewness (g) and kurtosis (h) parameters using two nonlinear transformations [1]. The formula can be defined as follows:

$$y = \frac{1}{g}\left[exp(gz-1)exp\left(\frac{hz^2}{2}\right)\right] \qquad with\ g \neq 0, h\epsilon R \qquad (1)$$

where,

z is a standard normal distribution;

g represents skewness parameter; and

h represents kurtosis parameter. Because the equation (1) cannot solve as usual. So, we will use another formula as:

$$y = a + bz\left(\frac{exp(g^z) - 1}{g^z}\right)\exp\left(\frac{hz^2}{2}\right) \qquad (2)$$

where,

a represents the location parameter; and

b represents the scale parameter. If the distribution does not include kurtosis that means h=0, then it can be rewritten the equation (2) as follows:

$$y_g = a + bz\left(\frac{exp(g^z) - 1}{g^z}\right) \qquad (3)$$

The equation (3) is called g-distribution that means the distribution has only skewness property. On the other hand, if the distribution does not have skewness property (symmetry), that means g=0, then it can be rewritten the equation (2) as follows:

$$y_h = a + bz\ exp\left(\frac{hz^2}{2}\right) \qquad (4)$$

The equation (4) is called h-distribution that means the distribution has only kurtosis property.

We shall explain some statistical properties of the Tukey gh distribution. First of all, this distribution is a continuous random variable. So, we can write the probability density function as [5, 9]:

$$f_x(x_p) = f_x(A + By_p) = \frac{1}{B}t_g, h(y_p), \qquad (5)$$

where,

x_p is the p^{th} quantile of x, $\quad 0.5 < p < 1$

y_p is the p^{th} quantile of y.

Furthermore, the cumulative distribution function of the Tukey gh distribution, F_y can be denoted as:

$$\int_a^b f_u\,(v)du = F_u(b) - F_u(a) \qquad (6)$$

where,

$a = T_{g,h}^{-1}(a)$ represents the inverse transformation of the Tukey gh distribution.

$b = T_{g,h}^{-1}(b)$ represents the inverse transformation of the Tukey gh distribution.

The equation (6) can be solved numerically because it is not explicit form.

3. The empirical Bayesian method

We will use the empirical Bayesian because the maximum likelihood function is not explicit formula. So, we will Approximated Bayesian computation (*ABC*) method. The main idea of ABC is to sample from approximated posterior distribution $\widehat{\mathrm{P}}(\theta|\mathrm{y}_1, \dots, \mathrm{y}_\mathrm{i})$, under some statistical conditions as shown in [5]. We will use algorithm of (*ABC*) as shown in [7], as follows:

- Draw a, b, g and h from the prior distribution $\pi(.)$;
- Generate y.
- Accept a, b, g and h with a normalized Markov kernel, otherwise re-sample our parameters again.

4.The parametric Bayesian method

As we have known that the parametric Bayesian method depends on the maximum likelihood function and prior distribution of the our parameters. So, the parametric Bayesian method can be written as:

$$\pi(y_n|y_1, y_2, \dots, y_{n-1}) = \pi(y_1, y_2, \dots, y_n) \times L(y_1, y_2, \dots, y_n) \qquad (7)$$

where,

$\pi(y_n|y_1, y_2, \dots, y_{n-1})$ is a posterior probability;

$\pi(y_1, y_2, \dots, y_n)$ is a prior probability; and

$L(y_1, y_2, \dots, y_n)$ is a maximum likelihood function.

As we mentioned before, the Tukey gh function can be solved numerically. So, the exactly maximum likelihood function does not exist. Therefore, the formula of likelihood function can be written as follows [1]:

$$L(\eta|y_1, y_2, \dots, y_n) = \Pi_{i=1}^{n} \left[\frac{f_z[T^{-1}(u_i;\theta)]}{\sigma_T[T^{-1}(u_i;\theta)]}\right] \quad (8)$$

Where $u_i = \frac{y_i-\mu}{\sigma}$

The equation (8) has solved approximately. Then, we will use the approximated likelihood function. We will use the parametric Bayesian estimating using Markov chain Monte Carlo (MCMC) to find the posterior distribution for a, b, g and h, using Metropolis Hasting (MH) Algorithm, where $\theta = [a, b, g, h]$.

The (MH) Algorithm is the discrete stochastic process by accepting-rejecting method to arrive for the target distribution [2]. So, the (MH) Algorithm is:

- Draw $\theta \sim \pi(\theta)$
- Set $i = 1,2 \dots,$
  - Sampling θ^* from the proposal distribution, $i.e., \pi(\theta^*|\theta^{i-1})$
  - Compute the ratio:

    $$r_g = \frac{\pi(\theta^*|y)}{\pi(\theta^{i-1}|y)},$$

  - Set

$$\theta^* = \begin{cases} \theta^* & with\ probability\ \min(1, r_\theta) \\ \theta^{i-1} & Otherwise \end{cases}$$

Now, we will apply our methods using simulation and real data.

5. Simulation study

We will apply our methods to estimate the parameters $\theta = [a, b, g, h]$ using Bayesian methods. We simulate the tukey gh distribution when $\theta = [a = 0.1, b = 0.9, g = 0.01, h]$ with different samples $N = [100, 125, 150]$ and iteration =10000, using R Programmed [8].

For The sample N= 100, we will estimate the parameters in the tukey gh distribution using ABC and MCMC methods when n= 100. Figure 1 presents the estimation of $\theta = [a, b, g, h]$ when N=100 using ABC method.

It is clear that we get the good trace plot of a,b, g and h whereas the bad trace plot of B. Figure 2 presents the estimation of $\theta = [a, b, g, h]$ when N= 100 using MCMC method. It is clear that we get the good estimation of all parameters.

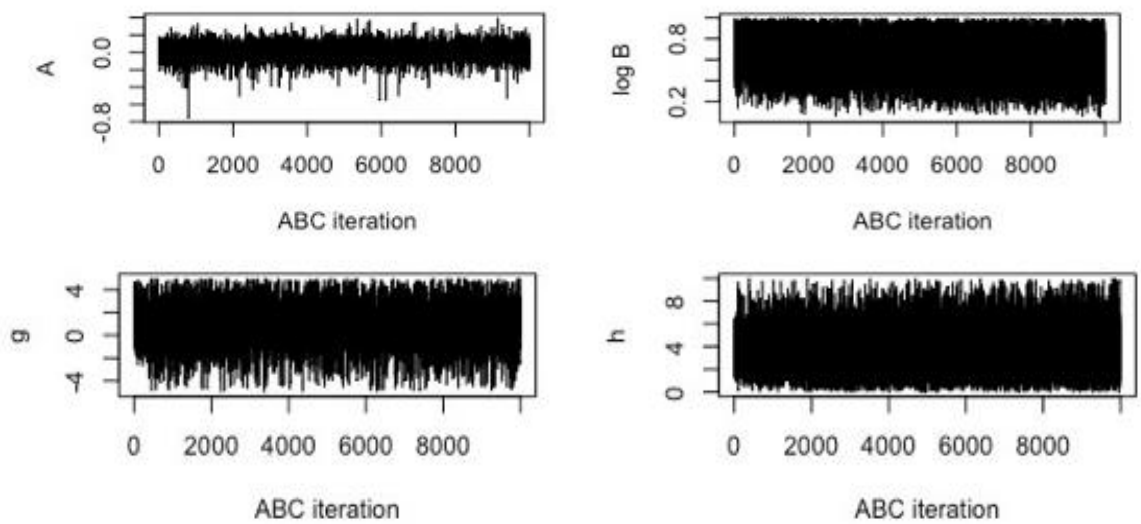

Figure 1. The posterior distribution of all parameters through the trace plot using ABC method when N=100

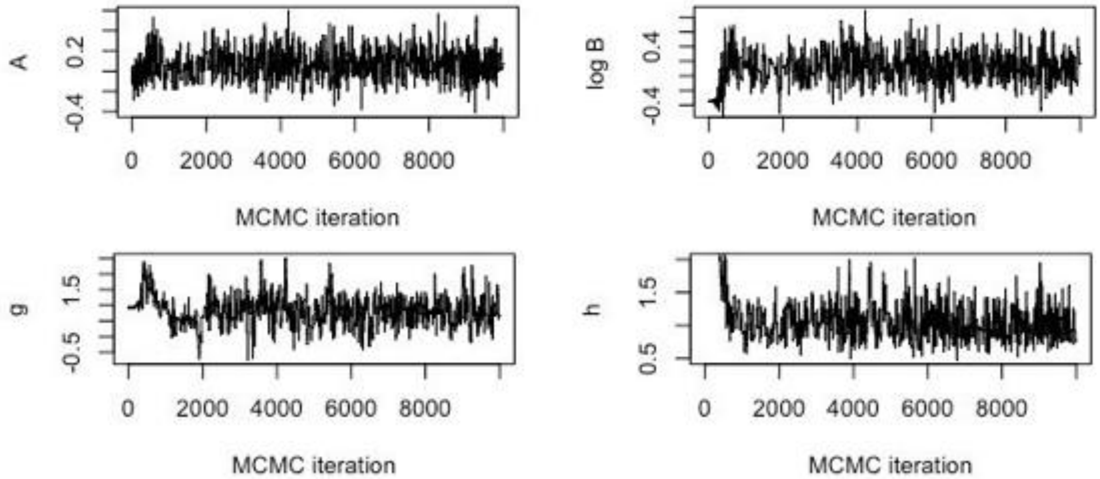

Figure 2. The posterior distribution of all parameters through the trace plot using MCMC method when N=100

By the same way, we estimate the parameters $\theta = [a, b, g, h]$ when N=125 using ABC and MCMC methods. Figure 3 shows the estimation of parameter $\theta = [a, b, g, h]$ when N= 125 using ABC method. In general it is obviously that the good estimation for all parameters, particularly A and g.

Figure 4 presents the estimation of parameters $\theta = [a, b, g, h]$ when N= 125 using MCMC method. It is clear that the good estimation for all parameters.

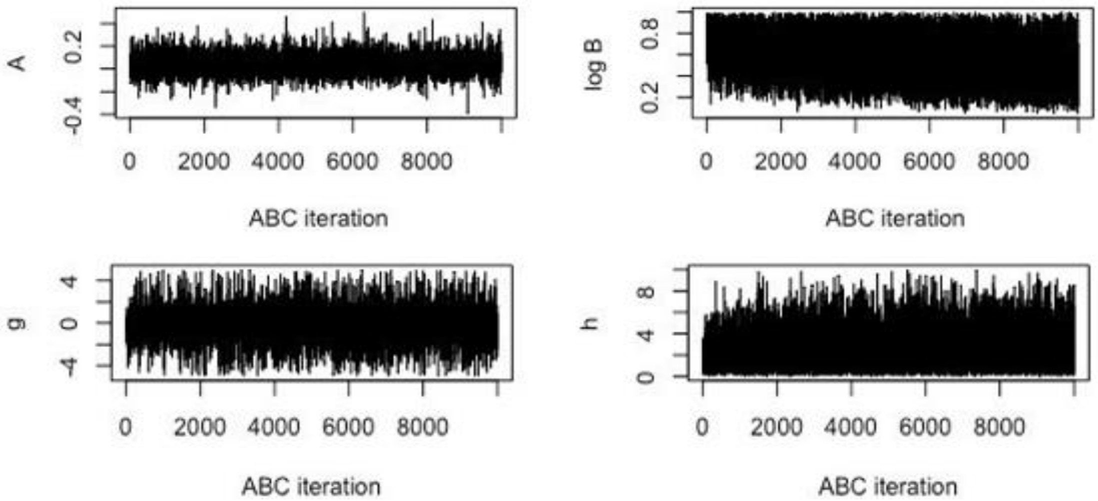

Figure 3. The posterior distribution of all parameters through the trace plot using ABC method when N=125

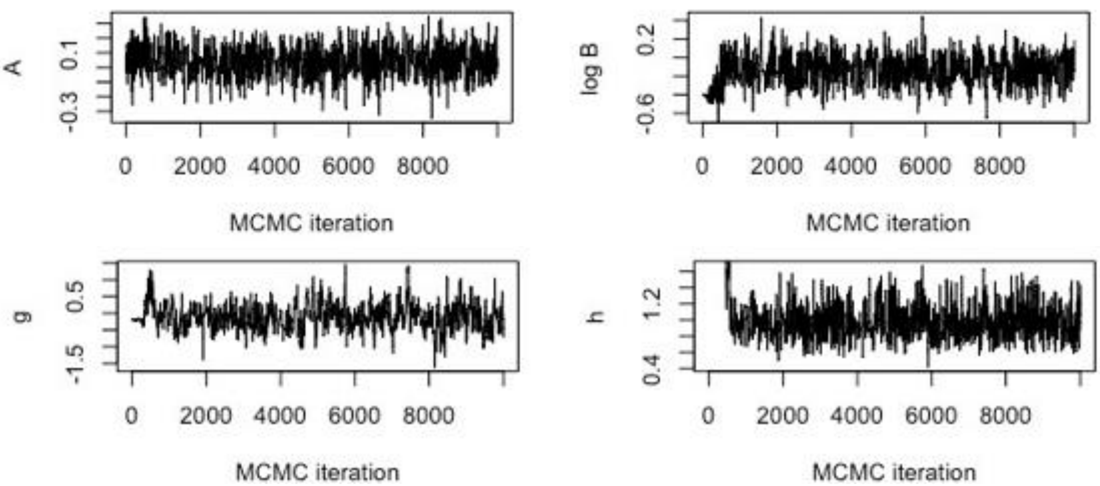

Figure 4. The posterior distribution of all parameters through the trace plot using MCMC method when N=125

Now, we apply ABC and MCMC method for estimating $\theta = [a, b, g, h]$ when N= 150. Figure 5 presents the estimation of parameters $\theta = [a, b, g, h]$ when N= 150 using ABC. It is a good estimation for A, g and rather h.
Figure 6 shows the estimation of parameters $\theta = [a, b, g, h]$ when N= 150 using MCMC. It is a good estimation for all parameters.

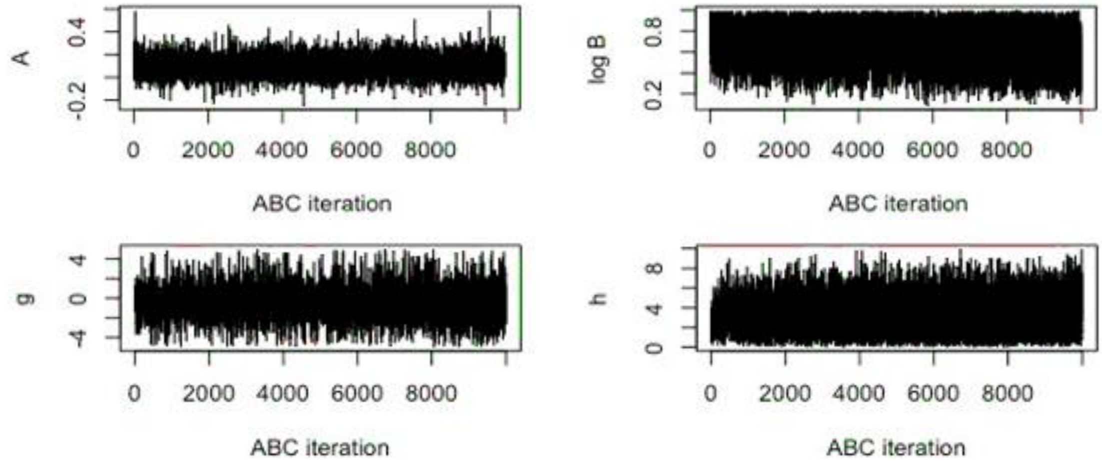

Figure 5. The posterior distribution of all parameters through the trace plot using ABC method when N=150

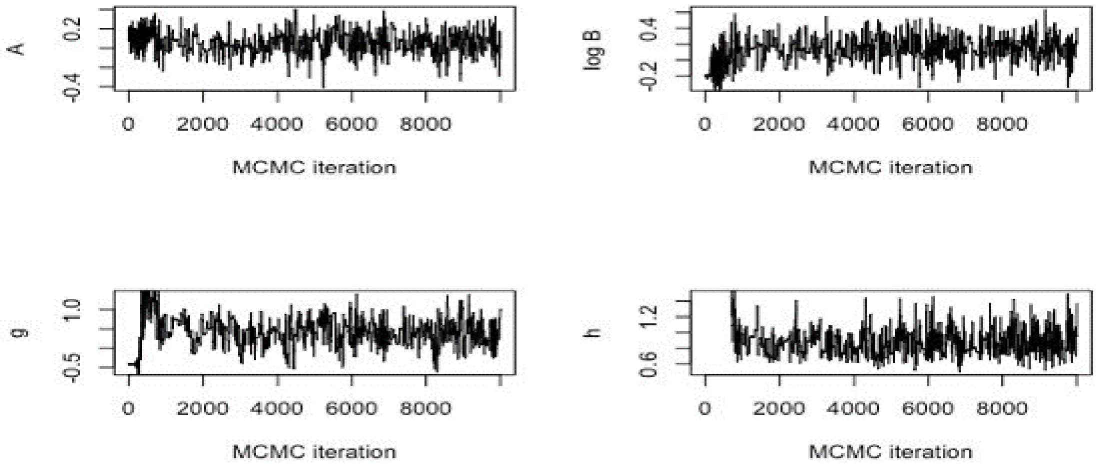

Figure 6. The posterior distribution of all parameters through the trace plot using MCMC method when N=150

6. Real data

We have adopted the index of stock market in Iraq, which is called (ISX60). We get the real data from 2017-2019 daily data. Figure 7 shows the daily data for ISX60. We will use the rate of return for ISX60. We applied ABC and MCMC methods for ISX60. Figure 8 presents the estimation of parameter a, b, g and h, using ABC method. It is clear that the results for our data are bad estimations for all parameters. Figure 9 shows the estimation of parameters a, b, g, h using MCMC method. It is clear that the good estimation for all parameters using MCMC method.

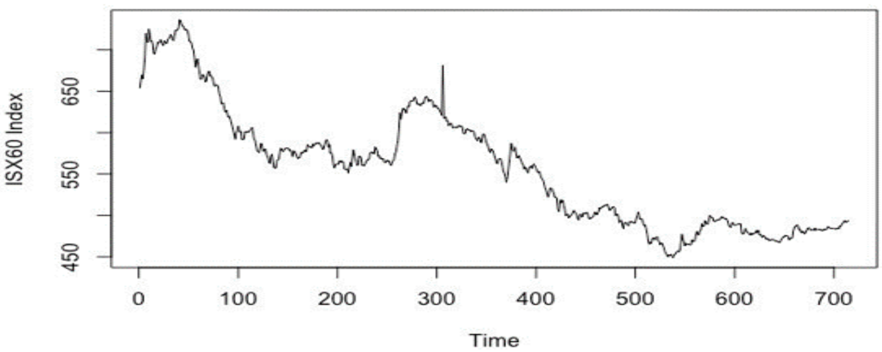

Figure 7. The movement of daily ISX60 index for 2017-2019

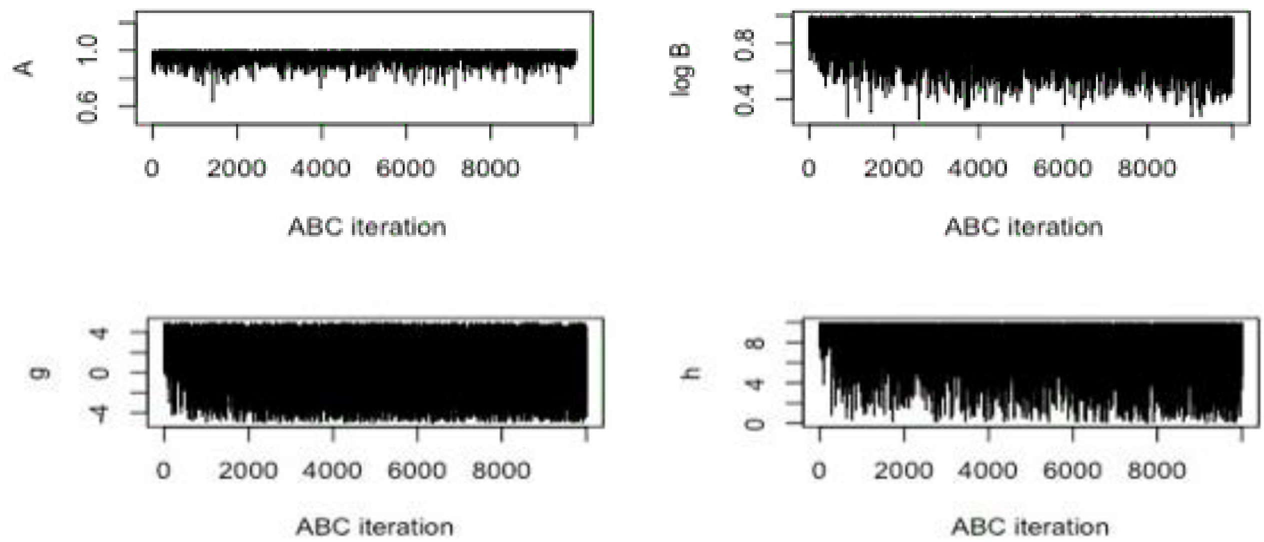

Figure 8. The posterior distribution of all parameters through the trace plot using ABC method for ISX60

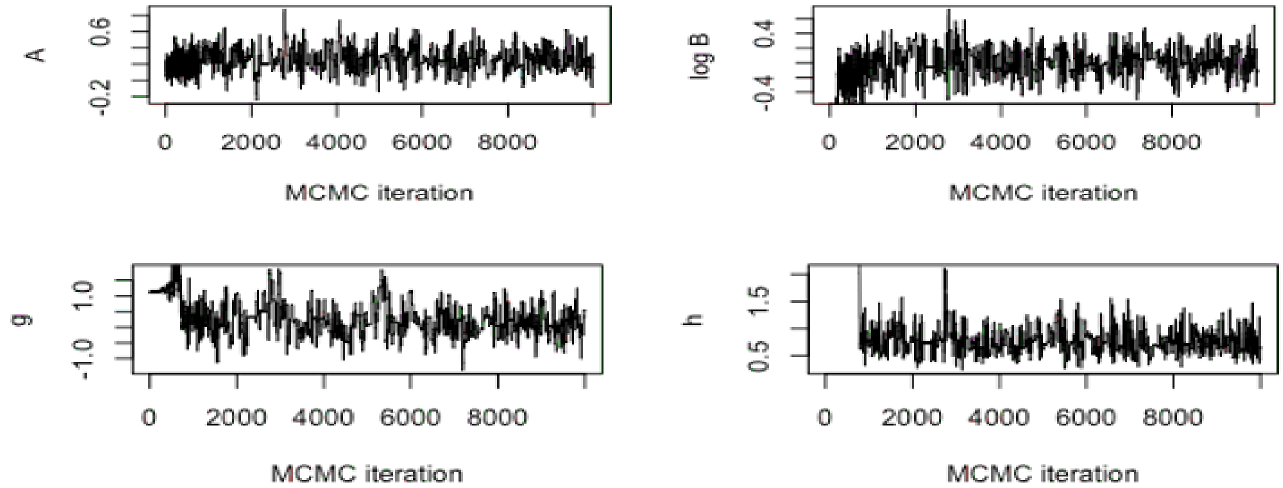

Figure 9. The posterior distribution of all parameters through the trace plot using MCMC method for ISX60

We have used the index of standard and Poor's in USA, which is called (SP500). We get the real data 2022 daily data. Figure 10 shows the daily data for SP500. We will use the rate of return for SP500. We applied ABC and MCMC methods for SP500. Figure 11 presents the estimation of parameter a, b, g and h, using ABC method. It is clear that the results for our data are bad estimations for all parameters. Figure 12 shows the estimation of parameters a, b, g, h using MCMC method. It is clear that the good estimation for all parameters using MCMC method.

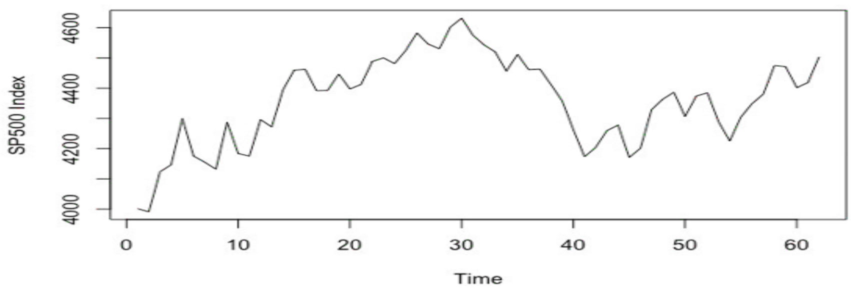

Figure 10. The movement of daily SP500 index for 2022

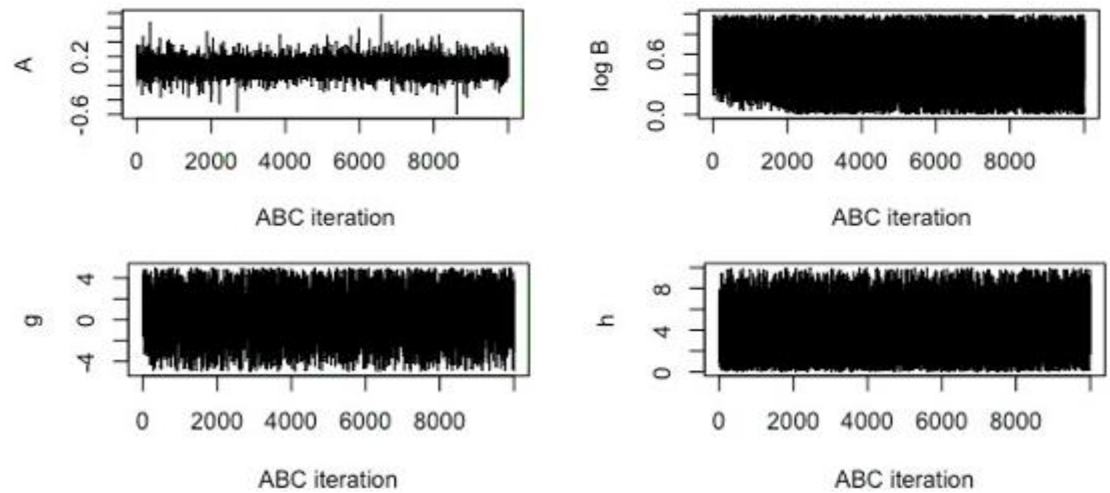

Figure 11. The posterior distribution of all parameters through the trace plot using ABC method for SP500

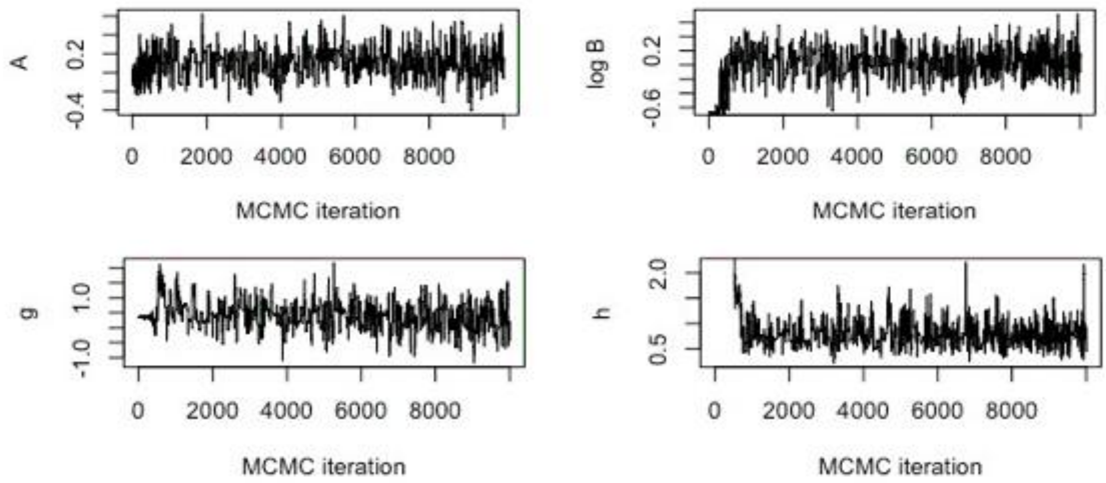

Figure 12. The posterior distribution of all parameters through the trace plot using MCMC method for SP500

7. Conclusion and further work

In this paper, we have applied some Bayesian methods for the Tukey gh distribution with different sampling size (n= 100, 125, 150), and then applied for real data such as ISX60 and SP500 index. We can conclude that the good method of estimating all parameters in simulation and real data is MCMC. I plan to apply the nonparametric Bayesian method to estimate all parameters because the likelihood function is very difficult to find it.

Declaration of competing interest

The authors declare that they have no any known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

1. J. Tukey, *Modern techniques in data analysis. NSF-Sponsored regional research conference at southeast on Massachusetts University*. Dartmouth, MA, 1977.

2. M. Al-Saadony, *Bayesian Stochastic Differential Equation Modeling with application*, Plymouth University, UK, 2013.
3. M. Al-Saadony, K. Bahr, and R. D. Tahir, “NON-BAYESIAN AND BAYESIAN METHODS TO ESTIMATE FRACTIONAL AR(1),” *Int. J. Agricult. Stat. Sci*, vol. 17, no. 1, pp. 993–1002, 2021.
4. G. Xu and M. G. Genton, “Efficient maximum approximated likelihood inference for Tukey’s g-and-h distribution,” Computational Statistics & Data Analysis, vol. 91, pp. 78–91, 2015.
5. M. Bee and L. Trapin, “A simple approach to the estimation of Tukey’s gh distribution,” *J. Stat. Comput. Simul.*, vol. 86, no. 16, pp. 3287–3302, 2016.
6. L. Möstel, M. Fischer, F. Pfälzner, and M. Pfeuffer, “Parameter estimation of Tukey-type distributions: A comparative analysis,” *Commun. Stat. Simul. Comput.*, vol. 50, no. 4, pp. 957–992, 2021.
7. P. Jk, M. T. Seielstatced, and A. Perez- Lezaun, “Population growth of Human y chromosomes: a study of y chromosome microsatellites,” *Mol Biol Evol*, vol. 16, pp. 1791–1798, 1999.
8. R. C. Team, *R: A Language and environment for statistical computing. R. Foundation for Statistical Computing*. Vienna, Austria, 2021.
9. Y. Yan and M. Genton, “The Tukey g. and h. distribution, significance,” Royal Statistical Society , pp. 12–13, 2019.