Sigma J Eng Nat Sci, Vol. 42, No. 5, pp. 1480−1489, October, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00114

Research Article

A heteroscedastic regression model with the generalized normal distribution

Emine Nur ESKİN[1], Fatma Zehra DOĞRU[1,*]

[1]*Department of Statistics, Giresun University, Giresun, 28100, Türkiye*

ARTICLE INFO

Article history
Received: 15 May 2023
Accepted: 12 July 2023

Keywords:
GN; Joint Location And Scale Model; Laplace Distribution; ML; Normal Distribution

ABSTRACT

In regression analysis, joint modeling mean and dispersion is an essential tool in absence of the variance homogeneity. Moreover, it is known in the literature that the generalized normal (GN) distribution has some features that provide flexibility in modeling thanks to its shape parameter. This paper proposes a joint location and scale model of the GN distribution for modeling location and scale in the presence of heteroscedasticity. We provide maximum likelihood (ML) estimators for the parameters of the proposed model. We also give an estimation procedure to estimate all parameters simultaneously. For the application, some simulation study scenarios and a real-life example are carried out to prove the estimation performance of the proposed model.

Cite this article as: Eskin EN, Doğru FZ. A heteroscedastic regression model with the generalized normal distribution. Sigma J Eng Nat Sci 2024;42(5):1480−1489.

INTRODUCTION

The normal distribution is the most widely recognised distribution because it has a wider range of applications in many fields. However, it cannot be a tractable model in case of deviation from normality. In such cases, it may be more appropriate to model, for example, a speech signal with a Laplacxe distribution. Therefore, the generalized normal (GN) distribution or the generalized Gauss distribution, having a wider application area, was proposed to solve this modeling problem. Moreover, the GN is a generalization of the Laplace, normal and uniform distributions. This enables GN distribution to be applied for modeling different types of data sets. It proves its usefulness in many sciences and engineering application areas such as adding watermarks to images [1], modeling speech signals [2], atmospheric noise, subband coding of audio and video signals [3], impulsive noise, the direction of arrival, modeling of the independent component analysis [4], blind signal separation [5] and so on. The GN distribution is the bell-shaped curve and its probability density function (pdf) has the peaky shape of the maximum. Furthermore, the distribution is symmetric to the location and the pdfs can be formed differently by shape and scale parameters [6]. This indicates that the pdf of GN distribution enables heavy-tailedness in a wide limit. It means that this family of distributions may have a thicker or thinner tail thickness than the normal distribution.

On the other hand, joint mean and dispersion models have been very important instruments for modeling heteroscedastic data sets in homogenous populations. There

***Corresponding author.**
*E-mail address: fatma.dogru@giresun.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

is extensive literature on joint mean and dispersion models which are summarized by [7] and [8]. This study also offers a chronicle summary of these studies in the following to locate their contribution to the existing literature.

Firstly, [9] offered a log-linear transformation on variance parameter to handle heteroscedasticity. Following him, the likelihood ratio test when heteroscedasticity exists was introduced by [10]. Later, modeling the variance heterogeneity for the normal regression was proposed by [11]. The parameters of the normal regression were estimated by [12] thanks to the restricted ML when the variances are dependent on log-linearly. By adopting generalized linear modeling, [13] explored an extension of the response surface approach to Taguchi-type robust designs. Subsequently, parameter estimation of the joint location and scale model of the t distribution was suggested by [14]. Joint modeling of mean and scale parameters, when the data set is longitudinal, was offered using a robust approach by [15]. In 2011, the same researchers, [16], examined Bayesian inference for joint modeling of location and scale parameters of the t distribution when the data set includes longitudinal observations. The joint modeling of the location and scale parameters of the SN distribution was examined by [17]. Recently, [7] investigated the joint modeling of location, scale, and skewness parameters of the skew Laplace normal (SLN) distribution, and [8] proposed joint location, scale, and skewness models with the mixtures of SLN distributions. Following them, some authors are interested in variable selection for the joint location and scale models. For instance, [18] considered variable selection for the inverse Gamma distribution and [19] investigated variable selection for the lognormal distribution. The variable selection for the joint location and scale model of the skew-normal (SN) [20,21] distribution was introduced by [22] and then followingly, [23] suggested variable selection for the parameters of joint location and scale model of the skew student-t-normal (STN) distribution. After that, [24] examined the variable selection for student-t regression models. [25] introduced variable selection in the joint location, scale, and skewness models of the SN distribution, and [26] proposed variable selection in the joint location, scale, and skewness models of the STN distribution.

The previous studies analyse the various versions of the GN distribution that have been obtained by adding the shape parameter to the normal distribution. The version 1 family is called exponential power distribution or generalized error distribution, a parametric family of symmetrical distributions. The exponential power distribution was first defined by [27] and later redefined as the GN distribution by [28]. Parameter estimation of the GN distribution using various estimation methods has been examined by [29], [28], and [30]. However, the location and scale parameters of the GN distribution were not considered in joint modeling; this paper proposes parameter estimation for the joint location and scale models of the GN distribution introduced by [28], see also [31]. This model can also be used as an alternative model for the joint modeling location and scale parameters of the normal and Laplace distributions since the GN distribution is more flexible than these distributions for modeling data sets.

The design of the paper is considered as follows. Section 2 defines the GN distribution and provides some properties of this distribution. Section 3 offers the parameter estimation thanks to the ML estimation method for the joint location and scale models of the GN distribution. Section 4 is devoted to applications, including a comprehensive simulation study and a real-life example to demonstrate the applicability of the proposed model. Section 5 presents some conclusions related to this study.

Generalized Normal Distribution

We assume that the random variable Y has a GN distribution with location parameter μ, scale parameter σ, and the shape parameter s considered in [28]. Its pdf is given below:

$$f(y) = \frac{s}{2\sigma\Gamma(1/s)} exp\left\{-\left|\frac{y-\mu}{\sigma}\right|^{s}\right\}, y \in \mathbb{R}, \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^{+}, s \in \mathbb{R}^{+}. \tag{1}$$

Here, $\Gamma(\cdot)$ denotes the gamma function. We note that the pdf given in (1) has two special cases according to s values. If $s = 1$ in equation (1), the distribution will be a Laplace distribution. On the other hand, if s is equal to 2 in equation (1), a normal distribution can be obtained. We further display Figure 1 to demonstrate the different pdf plots of the GN distribution. It can be observed from Figure 1 that the distribution is leptokurtic and heavy-tailed for small values of shape parameter s.

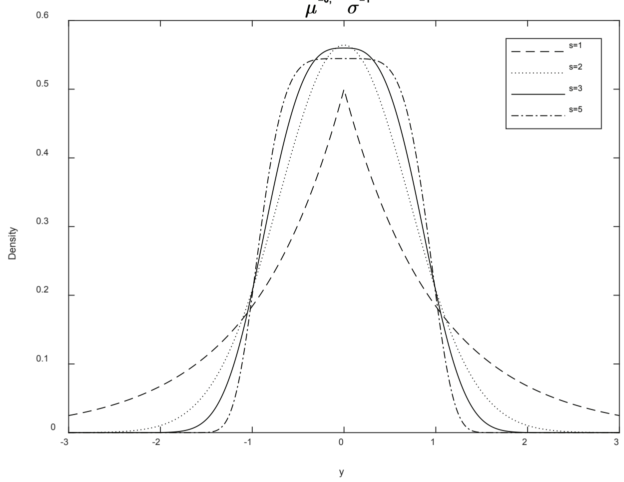

Figure 1. The pdf examples of the GN distribution for μ =0, σ = 1, and different parameter values of s.

The cumulative distribution function (cdf) of GN distribution presented by [28] has the following form:

$$F(y) = \frac{\Gamma\left(1/s, ((\mu - y)/\sigma)^s\right)}{2\Gamma(1/s)}, for\ y \leq \mu \tag{2}$$

$$F(y) = 1 - \frac{\Gamma\left(1/s, ((\mu - y)/\sigma)^s\right)}{2\Gamma(1/s)}, for\ y > \mu, \tag{3}$$

where $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1} \exp(-t)\, dt$ is the incomplete gamma function. We can give the following the n^{th} moment of the GN distribution about the origin for each positive n integer introduced by [28] to investigate some distributional measures of the GN distribution:

$$E(Y^n) = \frac{\mu^n \sum_{k=0}^{n} \binom{n}{k} (\sigma/\mu)^k \{1 + (-1)^k\} \Gamma((k+1)/s)}{2\Gamma(1/s)}. \tag{4}$$

We can obtain the expectation, variance, kurtosis, and skewness of the GN distribution as follows thanks to the n^{th} moment of the GN distribution given in (4):

$$E(Y) = \mu, \qquad Var(Y) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)},$$
$$Skewness = 0, \qquad Kurtosis = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma^2\left(\frac{3}{s}\right)}. \tag{5}$$

We see that the centre of the distribution is equal to μ and the skewness is zero. Besides, since variance and kurtosis depend on values of s, these measures become smaller concerning the shape parameter s.

Joint Location and Scale Models of the GN Distribution

Let us define the following joint location and scale models of the GN distribution:

$$\begin{cases} y_i \sim GN(\mu_i, \sigma_i, s), \quad i = 1,2,\dots,n \\ \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}, \\ \log \emptyset_i = \boldsymbol{z}_i^T \boldsymbol{\gamma}. \end{cases} \tag{6}$$

where $Var(y_i) = \frac{\sigma_i^2 \Gamma(3/s)}{\Gamma(1/s)} = \emptyset_i$, $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$ is the vector of observed responses, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^T$ and $\boldsymbol{z}_i = (z_{i1}, \dots, z_{iq})^T$ are the observed covariates corresponding to yi, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the $p \times 1$ vector of unknown parameters in the location model, and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ is the $q \times 1$ vector of unknown parameters in the scale model. Here, $\boldsymbol{z}_i$ may consist of some or all variables in $\boldsymbol{x}_i$, and other variables not involved in $\boldsymbol{x}_i$; so, the location and scale models may include different covariates, or some of the same covariates, and belong to common covariates differently. Let $\boldsymbol{x} = (x_1, \dots, x_n)^T$ and $\boldsymbol{z} = (z_1, \dots, z_n)^T$ be the covariate matrices.

Parameter Estimation of Joint Location and Scale Models of the GN Distribution

This part is devoted to finding ML estimators of parameters of joint location and scale models of the GN distribution. Consider the random sample $(y_i, x_i, \boldsymbol{z}_i)$, $i = 1,2,\dots, n$ from the model given in (6), and the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T, s)^T$. Maximization of the following log-likelihood function of the joint location and scale models of the GN distribution gives the $\widehat{\boldsymbol{\theta}}_{ML}$:

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, s) = \sum_{i=1}^{n} \left\{ \log s - \log 2 - \log c - \frac{\boldsymbol{z}_i^T \boldsymbol{\gamma}}{2} - \log(\Gamma(1/s)) \right\} - \sum_{i=1}^{n} \left| \frac{y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}}{c e^{\frac{z_i^T \gamma}{2}}} \right|^s, \tag{7}$$

where $\sigma = c e^{\frac{z_i^T \gamma}{2}}$, $c = \sqrt{\frac{\Gamma(1/s)}{\Gamma(3/s)}}$. For the estimation process, we obtain the score function vector

$$U(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(U_1^T(\boldsymbol{\beta}), U_2^T(\boldsymbol{\gamma})\right)^T, \tag{8}$$

where

$$U_1(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} s\, |y_i - x_i^T \boldsymbol{\beta}|^{s-1} sign(y_i - x_i^T \boldsymbol{\beta}) \frac{\boldsymbol{x}_i}{\left(c e^{\frac{z_i^T \gamma}{2}}\right)^s}, \tag{9}$$

$$U_2(\boldsymbol{\gamma}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} = -\sum_{i=1}^{n} \frac{z_i}{2} + \sum_{i=1}^{n} s|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}|^s \frac{\boldsymbol{z}_i}{2\left(c e^{\frac{z_i^T \gamma}{2}}\right)^{s+1}}. \tag{10}$$

and information matrix

$$H(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \end{bmatrix}, \tag{11}$$

where

$$\frac{\partial^2 \ell(\theta)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \frac{1}{\left(c e^{\frac{z_i^T \gamma}{2}}\right)^s} [2s|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}|^{s-1} \delta(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) + s(s-1)|y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}|^{s-2} sign(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2] \boldsymbol{x}_i \boldsymbol{x}_i^T, \tag{12}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} = -\sum_{i=1}^{n} \frac{s^2 |y_i - x_i^T \boldsymbol{\beta}|^{s-1}}{2\left(c e^{\frac{z_i^T \gamma}{2}}\right)^s} sign(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) \boldsymbol{x}_i \boldsymbol{z}_i^T, \tag{13}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \frac{s^2 |y_i - x_i^T \boldsymbol{\beta}|^{s-1}}{2\left(c e^{\frac{z_i^T \gamma}{2}}\right)^s} sign(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})\, \boldsymbol{z}_i \boldsymbol{x}_i^T, \tag{14}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} = -\sum_{i=1}^{n} \frac{s^2 |y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}|^s}{4\left(ce^{\frac{\boldsymbol{z}_i^T \boldsymbol{\gamma}}{2}}\right)^s} \boldsymbol{z}_i \boldsymbol{z}_i^T. \quad (15)$$

Here, $\delta(\cdot)$ is the delta Dirac function.

The following algorithm will allow us to find desired ML estimators for $\boldsymbol{\theta}$. In this algorithm, first, the estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are found and then the estimator for the parameter s is obtained using these estimators.

Algorithm:

Step 1. Set the starting values as $\boldsymbol{\theta}^{(0)} = \left((\boldsymbol{\beta}^{(0)})^T, \left(\boldsymbol{\gamma}^{(0)}\right)^T, s^{(0)}\right)^T$..

Step 2. Use $\boldsymbol{\beta}^{(k)}$, $\boldsymbol{\gamma}^{(k)}$ where $\boldsymbol{\theta}^{*(k)} = \left(\left(\boldsymbol{\beta}^{(k)}\right)^T, \left(\boldsymbol{\gamma}^{(k)}\right)^T\right)^T$; and compute the following $(k + 1)$th estimates:

$$\widehat{\boldsymbol{\theta}}^{*(k+1)} = \widehat{\boldsymbol{\theta}}^{*(k)} + \left[-H(\widehat{\boldsymbol{\theta}}^{*(k)})\right]^{-1} U\left(\widehat{\boldsymbol{\theta}}^{*(k)}\right). \quad (16)$$

Using the calculated $\widehat{\boldsymbol{\theta}}^{*(k+1)}$ find $\hat{s}^{(k+1)}$ that maximizes the following equation:

$$\frac{n}{s}\left(\frac{1}{s}\psi\left(\frac{1}{s}\right) + 1\right) - \sum_{i=1}^{n} \left|\frac{y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}^{(k+1)}}{ce^{\frac{\boldsymbol{z}_i^T \boldsymbol{\gamma}^{(k+1)}}{2}}}\right|^s \log\left(\left|\frac{y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}^{(k+1)}}{ce^{\frac{\boldsymbol{z}_i^T \boldsymbol{\gamma}^{(k+1)}}{2}}}\right|\right) = 0. \quad (17)$$

Step 3. Repeat these 2 steps until convergence is achieved.

NUMERICAL STUDIES

This section provides some simulation examples and a real data example to show the modeling performance of the joint location and scale models of the GN distributions. The computational details for the applications can be summarized as follows.

Notes on Implementation

i) All numerical studies including simulation and real data examples were performed using the MATLAB R2017b software.

ii) The stopping rule is taken as 10^{-6} for all numerical examples.

iii) The starting points for the proposed algorithm given in Section 3 are set as the actual parameter values as the starting points to carry out the estimation procedure in the simulation study. Further, estimation results of the heteroscedastic normal regression model by [11] were taken as initial values for all of the location and scale models in the real data example.

iv) Random sample generation from the GN distribution can be summarized in the following steps. We note that this generating procedure was proposed by [32].

- Sample X from the gamma distribution with parameters 1/s and γ:

$$X \sim Gamma(1/s, 1). \quad (18)$$

-Generate a random number from the independent random variable Z that takes the values -1 and +1 with 0.5 probability:

$$Z \sim \frac{1}{2}[Z = -1] + \frac{1}{2}[Z = 1]. \quad (19)$$

- Generate random numbers from the GN distribution with parameters μ, σ, and s using the following relationship:

$$Y = \mu + \sigma Z |X|^{1/s} \sim GN(\mu, \sigma, s). \quad (20)$$

Simulation Examples

This section offers two simulation scenarios to investigate the estimation performance of the joint location and scale models of the GN distributions. These two simulation scenarios include the cases when the parameter s is known and when it is unknown. The performance of the estimators is measured in terms of the bias and the mean squared error (MSE). The formulas for these measures are given by:

$$\widehat{bias}(\hat{\theta}) = \bar{\theta} - \theta, \;\; \widehat{MSE}(\hat{\theta}) = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_j - \theta\right)^2, \quad (21)$$

where θ represents the true parameter value, $\hat{\theta}_j$ shows the estimate of θ for the jth simulated data and $\bar{\theta} = \frac{1}{N}\sum_{j=1}^{N} \hat{\theta}_j$. The simulation studies are replicated $N = 1000$ times. The sample sizes (n) are considered 50, 100, 150, and 200 for all simulation scenarios.

The random sample generation is done by using the following location and scale models of the GN distribution

$$\begin{cases} y_i \sim GN(\mu_i, \sigma_i, s), \;\; i = 1,2,\dots,n \\ \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}, \\ \log \phi_i = \boldsymbol{z}_i^T \boldsymbol{\gamma}, \end{cases}$$

where all covariate vectors $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are generated from uniform distribution *Uniform*(-1,1), which are independent. For all simulation scenarios, the true parameter values are arbitrarily taken as $\boldsymbol{\beta} = (1,0.7,0.5)$ and $\boldsymbol{\gamma} = (-0.5, 0.3,0.2)$.

Scenario 1. This simulation scenario was conducted when the shape parameter s is known and simulation results are presented in Tables 1-4. These tables provide bias and MSE values of estimates for the sample sizes 50, 100, 150, and 200 and s = 1, 2, 3 and 5. We remind again that the GN distribution reduces to the Laplace distribution if s = 1 and reduces to the normal distribution if s = 2. We also include results for different parameter values of s to examine the flexibility of the GN distribution. The simulation results given in the tables can be summarized as follows: The proposed parameter estimation algorithm for parameter estimation works well to obtain parameter estimates. Moreover, the MSE values decrease as the sample size increases, i.e. the parameter estimates are consistent. According to bias values, we observe that estimates are close to true parameter values as the sample size increases.

Table 1. Values of bias and MSE for estimates (Scenario 1 when $s = 1$)

Model $s = 1$	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	-0.0019	0.0112	-0.0006	0.0046	-0.0034	0.0031	0.0003	0.0020
	β_2	0.0071	0.0385	-0.0023	0.0139	-0.0007	0.0092	-0.0001	0.0063
	β_3	0.0093	0.0361	-0.0014	0.0137	0.0013	0.0088	-0.0010	0.0059
Scale	γ_1	-0.0632	0.0206	-0.0303	0.0086	-0.0189	0.0051	-0.0181	0.0040
	γ_2	0.0334	0.0746	0.0105	0.0255	0.0108	0.0158	0.0040	0.0117
	γ_3	0.0215	0.0646	0.0048	0.0256	0.0048	0.0163	-0.0037	0.0118

Table 2. Values of bias and MSE for estimates (Scenario 1 when $s = 2$)

Model $s = 2$	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	-0.0024	0.0111	-0.0014	0.0052	-0.0030	0.0028	-0.0016	0.0023
	β_2	-0.0025	0.0335	-0.0004	0.0143	-0.0040	0.0096	0.0025	0.007
	β_3	-0.0056	0.0374	0.0023	0.0144	-0.0007	0.0082	-0.0029	0.0067
Scale	γ_1	-0.1904	0.0711	-0.1424	0.0369	-0.1258	0.0268	-0.1060	0.0192
	γ_2	-0.1356	0.172	-0.1240	0.0657	-0.1241	0.0467	-0.1265	0.0388
	γ_3	-0.0801	0.1501	-0.0854	0.0575	-0.0938	0.0386	-0.0820	0.0291

Table 3. Values of bias and MSE for estimates (Scenario 1 when $s = 3$)

Model $s = 3$	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	-0.0017	0.0120	0.0060	0.0039	-0.0003	0.0026	-0.0024	0.0019
	β_2	-0.0032	0.0360	-0.0011	0.0122	0.0000	0.0085	-0.0006	0.0061
	β_3	-0.0092	0.0326	0.0045	0.0116	0.0002	0.0072	-0.0039	0.0059
Scale	γ_1	-0.2106	0.0840	-0.2610	0.0793	-0.2441	0.0669	-0.2338	0.0601
	γ_2	-0.0883	0.1448	-0.1250	0.0555	-0.1365	0.0413	-0.1174	0.0312
	γ_3	-0.0908	0.1630	-0.0788	0.0424	-0.0869	0.0311	-0.0885	0.0254

Table 4. Values of bias and MSE for estimates (Scenario 1 when $s = 5$)

Model $s = 5$	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	-0.0155	0.0456	-0.0011	0.0080	0.0013	0.0020	-0.0016	0.0012
	β_2	0.0205	0.1822	-0.0025	0.0487	0.0021	0.0090	0.0010	0.0033
	β_3	-0.0055	0.1030	0.0058	0.0400	0.0017	0.0073	-0.0003	0.0036
Scale	γ_1	-0.1717	0.4134	-0.2554	0.2390	-0.2676	0.2236	-0.2818	0.0835
	γ_2	-0.1068	0.4007	-0.1186	0.1470	-0.1136	0.0626	-0.1202	0.0349
	γ_3	-0.0880	0.2802	-0.0785	0.1986	-0.0715	0.1789	-0.0882	0.0274

Table 5. Values of bias and MSE for estimates (Scenario 2 when s = 3)

Model s = 3	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	0.0138	0.0456	0.0039	0.0080	-0.0027	0.0020	-0.0010	0.0012
	β_2	-0.0003	0.1822	-0.0048	0.0487	0.0010	0.0090	0.0019	0.0033
	β_3	-0.0119	0.1030	-0.0030	0.0400	0.0004	0.0073	-0.0007	0.0036
Scale	γ_1	-0.3039	0.4134	-0.2660	0.2390	-0.2408	0.2236	-0.2269	0.0835
	γ_2	-0.1017	0.4007	-0.1345	0.1470	-0.1283	0.0626	-0.1334	0.0349
	γ_3	-0.0463	0.2802	-0.0802	0.1986	-0.0806	0.1789	-0.0899	0.0274
	s	0.2677	0.1136	0.3144	0.1127	0.3256	0.1124	0.3279	0.1118

Table 6. Values of bias and MSE for estimates (Scenario 2 when s = 5)

Model s = 5	n	50		100		150		200	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Location	β_1	-0.0048	0.0456	0.0060	0.0080	-0.0001	0.0020	-0.0001	0.0012
	β_2	-0.0277	0.1822	0.0064	0.0487	0.0007	0.0090	-0.0084	0.0033
	β_3	-0.0019	0.1030	-0.0052	0.0400	0.0045	0.0073	0.0033	0.0036
Scale	γ_1	-0.0516	0.4134	-0.1084	0.2390	-0.1497	0.2236	-0.2273	0.0835
	γ_2	-0.0805	0.4007	-0.0873	0.1470	-0.0813	0.0626	-0.1230	0.0349
	γ_3	-0.0944	0.2802	-0.0773	0.1986	-0.0874	0.1789	-0.0730	0.0274
	s	0.7400	3.0065	0.6307	2.7641	0.6569	2.7588	0.6403	2.7203

Scenario 2. This scenario provides parameter estimation when s is unknown. In this simulation example, we estimate all parameters simultaneously using the estimation algorithm given in sub-section 3.1. The estimation results are given in Tables 5 and 6 for different sample sizes (n = 50, 100, 150 and 200) and the parameters s = 3 and 5. These tables consist of values of bias and MSE for estimates. From the tables, we observe that the proposed algorithm works effectively to estimate all parameters simultaneously. Furthermore, MSE values become smaller as the sample size increases.

Martin Marietta Data Set

This part is devoted to a real-life example of the applicability of the proposed joint location and scale models of the GN distribution. The Martin Marietta data set consists of the excess rate of gains of the Marietta company and an index for the excess rate of gain for the New York Exchange (CRSP). For this data set, the company's gains and the CRSP index were evaluated monthly for five years. This data set was used by [33] for normal regression analysis. Furthermore, in the Martin Marietta data set, the dense and irregular observations observed in the tails offer a robust method, and therefore, [34] and [35] proposed to use the Martin Marietta data set to model the skew t regression model. This data set was investigated by [14] for parameter estimation of the joint location and scale model based on the t distribution. Besides, [7] analysed this data set to estimate parameters of the joint location, scale, and skewness model based on the skew Laplace normal distribution.

In the literature, this data set is mostly considered for the case of heteroscedasticity. In order to prove that the Martin Marietta data set does not satisfy the homoscedasticity assumption, we will apply the Breusch-Pagan test proposed by [36]. The hypotheses for this test are as follows:

H_0*: Homoscedasticity is present (the residuals are distributed with equal variance),*

H_1*: Heteroscedasticity is present (the residuals are not distributed with equal variance)*

The p-value of the Breusch-Pagan test is 2.36×10^{-11} which proves that heteroscedasticity is present. This result is also supported with the help of the plot of residuals versus fitted values given in Figure 2. Therefore, since the Martin Marietta data contains heteroscedasticity, we consider using this data set to demonstrate the modeling performance of the joint location and scale models of the GN distribution. We will also compare the proposed model with the joint location and scale models of the normal and Laplace distributions.

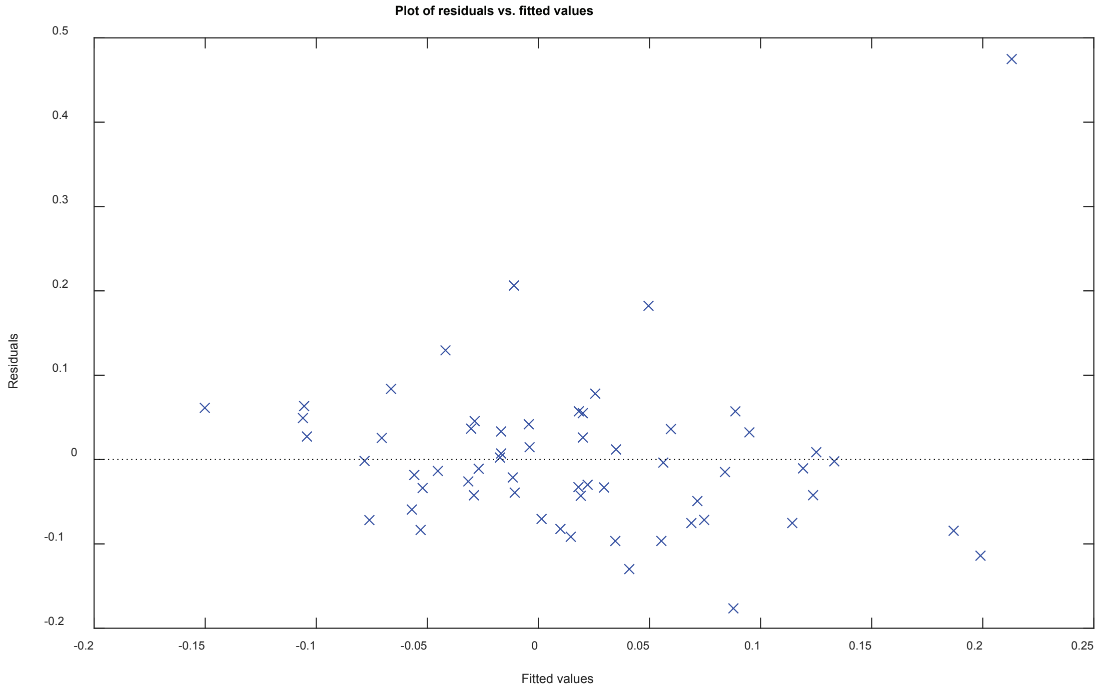

Figure 2. The plot of residuals versus fitted values.

We will use the following information criteria for comparing the models:

$$-2\ell(\hat{\boldsymbol{\theta}}) + mc_n\,, \tag{22}$$

where $l(\cdot)$ represents the maximized log-likelihood, m shows the number of estimated parameters, c_n indicates the penalty term, and n is the number of observations. For the penalty term, we set $c_n = 2$ for the Akaike Information Criterion (AIC) introduced by [37], $c_n = \log(n)$ for Bayesian Information Criterion (BIC) proposed by [38], $c_n = 0.2\sqrt{n}$ for Efficient Determination Criterion (EDC) defined by [39], and $c_n = \log n + 1$ for the Consistent Akaike Information Criterion (CAIC) proposed by [40].

Table 7 gives the estimation results, maximized likelihood, and the information criteria for the compared models. We note that the estimation procedure for the joint location and scale models of the Laplace distribution is obtained by modifying the estimation results for the Laplace regression model given by [41] and studied by [42]. The estimation procedure for the joint location and scale models of the Laplace distribution is summarized and

Table 7. Estimation results for the compared models using the real-life data set.

		GN	Normal	Laplace
		Estimate	**Estimate**	**Estimate**
Location model	β_0	0.0032	-0.0021	-0.0202
	β_1	1.3846	1.3121	1.2465
Scale model	γ_0	-5.3844	-5.3676	-2.1581
	γ_1	18.0275	18.4708	-14.1811
	s	2.1700	-	-
Information criteria	$\ell(\hat{\boldsymbol{\theta}})$	**137.4844**	126.1511	110.8426
	AIC	**-264.9689**	-244.3023	-213.6851
	CAIC	**-249.4971**	-231.9249	-201.3077
	BIC	**-254.4971**	-235.9249	-205.3077
	EDC	**-267.2229**	-246.1055	-215.4883

Note: The bold texts indicate the highest value of the maximized likelihood and the lowest values of the information criteria.

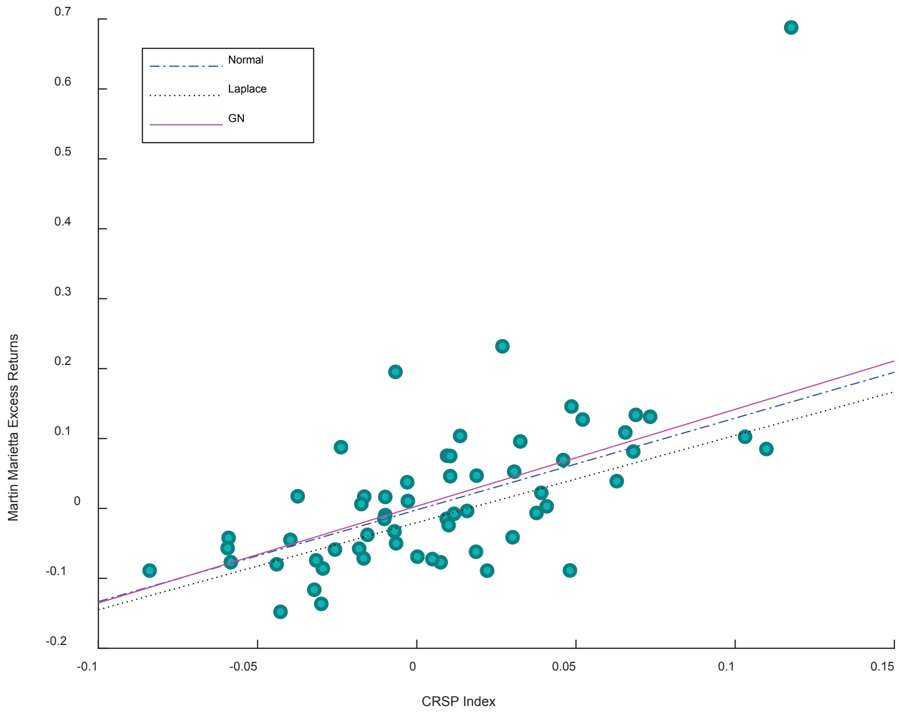

Figure 3. The scatterplot of the real-life data set as well as the fitted regression lines of joint location, and scale models of the normal, Laplace, and GN distributions.

presented in the Appendix. From Table 7 we observe that the maximum log-likelihood value and the smallest information criteria values are provided from the joint location and scale models of the GN distribution. This result is also supported by Figure 3, where the best fit is obtained by the joint location and scale models of the GN distribution.

CONCLUSION

This article offered parameter estimation of the joint location and scale parameters of the GN distribution, which is a helpful tool in modeling heteroscedasticity. This model is also more flexible than joint location and scale models of normal and Laplace distributions. We provided an estimation methodology for the joint location and scale model of the GN distribution and proposed an algorithm to estimate parameters of interest simultaneously. We performed some simulation examples for the cases where the shape parameter is known and unknown. From the simulation examples, we observed that the proposed algorithm works accurately in estimating all parameters. We also provided a real-life data set to demonstrate the modeling performance of the joint location and scale model of the GN distribution according to joint location and scale models of normal and Laplace distributions. The results of the real-life data set showed that the best model is obtained from the joint location and scale model of the GN distribution when compared with the other models in terms of information criteria. Finally, alternatively, the newly proposed GN joint location and scale model can be chosen to model heteroscedasticity over the normal and Laplace joint location and scale models.

ACKNOWLEDGEMENTS

We are grateful to the associate editor and reviewers for their insightful comments, which considerably improved the manuscript.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Briassouli A, Tsakalides P, Stouraitis A. Hidden messages in heavy-tails: DCT-domain watermark detection using alpha-stable models. IEEE Trans Multimedia 2005;7:700-715. [CrossRef]

[2] Kokkinakis K, Nandi AK. Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling. Signal Process 2005;85:1852-1858. [CrossRef]

[3] Sharifi K, Leon-Garcia A. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. IEEE Trans Circuits Syst Video Technol 1995;5:52-56. [CrossRef]

[4] Choi S, Cichocki A, Amari S-I. Flexible Independent Component Analysis. J VLSI Signal Process Syst Signal Image Video Technol 2000;26:25-38. [CrossRef]

[5] Wu H-C, Principe JC. Minimum entropy algorithms for source separation. 1998 Midwest Symposium on Circuits and Systems (Cat. No. 98CB36268), 1998. p. 242-245.

[6] Varanasi MK, Aazhang B. Parametric generalized Gaussian density estimation. J Acoust Soc Am 1989;86:1404-1415. [CrossRef]

[7] Doğru FZ, Arslan O. Joint modelling of the location, scale and skewness parameters of the skew laplace normal distribution. Iran J Sci Technol Trans A Sci 2019;43:1249-1257. [CrossRef]

[8] Doğru FZ, Yu K, Arslan O. Heteroscedastic and heavy-tailed regression with mixtures of skew Laplace normal distributions. J Stat Comput Simul 2019;89:3213–3240. [CrossRef]

[9] Park RE. Estimation with heteroscedastic error terms. Econometrica 1966;34:888. [CrossRef]

[10] Harvey AC. Estimating regression models with multiplicative heteroscedasticity. Econometrica 1976:461-465. [CrossRef]

[11] Aitkin M. Modelling variance heterogeneity in normal regression using GLIM. J R Stat Soc Ser C Appl Stat 1987;36:332-339. [CrossRef]

[12] Verbyla AP. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. Journal of the Royal Statistical Society: Series B (Methodological) 1993;55:493-508. [CrossRef]

[13] Engel J, Huele AF. A generalized linear modeling approach to robust design. Technometrics 1996;38:365-373. [CrossRef]

[14] Taylor J, Verbyla A. Joint modelling of location and scale parameters of the t distribution. Stat Modelling 2004;4:91-112. [CrossRef]

[15] Lin T-I, Wang Y-J. A robust approach to joint modeling of mean and scale covariance for longitudinal data. J Stat Plan Inference 2009;139:3013-3026. [CrossRef]

[16] Lin T-I, Wang W-L. Bayesian inference in joint modelling of location and scale parameters of the t distribution for longitudinal data. J Stat Plan Inference 2011;141:1543-1553. [CrossRef]

[17] Li H, Wu L. Joint modelling of location and scale parameters of the skew-normal distribution. Applied Mathematics-A Journal of Chinese Universities 2014;29:265-272. [CrossRef]

[18] Wu L, Li H. Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. Metrika 2012;75:795-808. [CrossRef]

[19] Zhang Z. Variable selection for joint mean and dispersion models of the lognormal distribution. Hacettepe J Math Stat 2012;41:307-320.

[20] Azzalini A. A class of distributions which includes the normal ones. Scand J Stat 1985;12:171-178.

[21] Azzalini A. Further results on a class of distributions which includes the normal ones. Statistica 1986;46:199-208.

[22] Wu L-C, Zhang Z-Z, Xu D-K. Variable selection in joint location and scale models of the skew-normal distribution. J Stat Comput Simul 2013;83:1266-1278. [CrossRef]

[23] Wu L-C. Variable selection in joint location and scale models of the skew-t-normal distribution. Commun Stat Simul Comput 2014;43:615-630. [CrossRef]

[24] Zhao W, Zhang R. Variable selection of varying dispersion student-t regression models. J Syst Sci Complex 2015;28:961-977. [CrossRef]

[25] Li H, Wu L, Ma T. Variable selection in joint location, scale and skewness models of the skew-normal distribution. J Syst Sci Complex 2017;30:694-709. [CrossRef]

[26] Wu L, Tian G-L, Zhang Y-Q, Ma T. Variable selection in joint location, scale and skewness models with a skew-t-normal distribution. Stat Interface 2017;10:217-227. [CrossRef]

[27] Subbotin MT. On the law of frequency of error. Matematicheskii Sbornik 1923;31:296-301.

[28] Nadarajah S. A generalized normal distribution. J Appl Stat 2005;32:685-694. [CrossRef]

[29] Varanasi MK, Aazhang B. Parametric generalized Gaussian density estimation. J Acoust Soc Am 1989;86:1404-1415. [CrossRef]

[30] Roenko AA, Lukin V V, Djurović I, Simeunović M. Estimation of parameters for generalized Gaussian distribution. 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), IEEE; 2014. p. 376-379. [CrossRef]

[31] Eskin EN. Joint modelling of the location and scale parameters of the generalized normal distribution (master's thesis). Giresun: Giresun University; 2022.

[32] Dominguez-Molina JA, González-Farías G, Rodríguez-Dagnino RM, Monterrey IC. A practical procedure to estimate the shape parameter in the generalized Gaussian distribution. Technique Report I-01-18_eng Pdf, Available through Http://Www Cimat Mx/Reportes/Enlinea/I-01-18_eng Pdf 2003;1.

[33] Butler RJ, McDonald JB, Nelson RD, White SB. Robust and partially adaptive estimation of regression models. Rev Econ Stat 1990;72:321-327. [CrossRef]

[34] Jones MC. A skew t distribution. Probability and Statistical Models with Applications. New York: Taylor & Francis Group; 2001. p. 269-278. [CrossRef]

[35] Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J R Stat Soc Series B Stat Methodol 2003;65:367-389. [CrossRef]

[36] Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1979;47:1287–1294. [CrossRef]

[37] Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike, New York, NY: Springer New York; 1998, p. 199-213. [CrossRef]

[38] Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics 1978;6:461-464. [CrossRef]

[39] Bai Z.-D, Krishnaiah PR, Zhao L.-C. On rates of convergence of efficient detection criteria in signal processing with white noise. IEEE Trans Inf Theory 1989;35:380-388. [CrossRef]

[40] Bozdogan H. Choosing the Number of Component Clusters in the Mixture-Model Using a New Informational Complexity Criterion of the Inverse-Fisher Information Matrix. In: Opitz O, Lausen B, Klar R, editors. Information and Classification, Berlin, Heidelberg: Springer Berlin Heidelberg; 1993, p. 40-54. [CrossRef]

[41] Phillips RF. Least absolute deviations estimation via the EM algorithm. Stat Comput 2002;12:281-285.

[42] Yang F. Robust mean change-point detecting through laplace linear regression using EM algorithm. J Appl Math 2014;2014:856350. [CrossRef]

[43] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B Stat Methodol 1977;39:1-22. [CrossRef]

APPENDIX

We assume that the following model has the form of joint location and scale models of the Laplace distribution:

$$\begin{cases} y_i \sim L(\mu_i, \sigma_i), \ \ i = 1,2,\dots,n \\ \mu_i = \boldsymbol{x}_i{}^T\boldsymbol{\beta}, \\ \log \sigma_i = \boldsymbol{z}_i{}^T\boldsymbol{\gamma}. \end{cases} \tag{23}$$

where $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$ shows the vector of observed responses, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^T$ and $\boldsymbol{z}_i = (z_{i1}, \dots, z_{iq})^T$ are the observed covariates, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is the parameter in the location model, and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ is the parameter in the scale model.

In this section, we will give the summarization of the estimation procedure with the help of the EM algorithm proposed by [43] for the joint location and scale models of the Laplace distribution given in (23). The steps are given with the following algorithm:

EM Algorithm:

1. Set the starting point for $\boldsymbol{\theta}$ as $\boldsymbol{\theta}^{(0)} = \left((\boldsymbol{\beta}^{(0)})^T, (\boldsymbol{\gamma}^{(0)})^T\right)^T$.

2. E Step: Using the given values calculate

$$\hat{\tau}_i^{(k)} = \frac{e^{\boldsymbol{z}_i{}^T\hat{\boldsymbol{\gamma}}^{(k)}}}{\left|y_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}^{(k)}\right|}$$

and obtain the following objective function:

$$Q_c = E(\ell_c(\boldsymbol{\beta},\boldsymbol{\gamma})|y_i) = -\sum_{i=1}^{n} \boldsymbol{z}_i{}^T\boldsymbol{\gamma} - \frac{1}{2}\sum_{i=1}^{n} \hat{\tau}_i^{(k)} \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2}{\left(e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}\right)}.$$

3. M Step: Maximize the objective function concerning parameters. Use the current values $\boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{\theta}^{(k)} = \left((\boldsymbol{\beta}^{(k)})^T, (\boldsymbol{\gamma}^{(k)})^T\right)^T$ and compute $\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \left[-H(\widehat{\boldsymbol{\theta}}^{(k)})\right]^{-1} U(\widehat{\boldsymbol{\theta}}^{(k)})$.

Here, the first derivatives of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be found as:

$$U_1(\boldsymbol{\beta}) = \frac{\partial Q_c}{\partial \boldsymbol{\beta}} = -\sum_{i=1}^{n} \hat{\tau}_i \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})}{e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}} \boldsymbol{x}_i\,, \tag{24}$$

$$U_2(\boldsymbol{\gamma}) = \frac{\partial Q_c}{\partial \boldsymbol{\gamma}} = -\sum_{i=1}^{n} \boldsymbol{z}_i + \sum_{i=1}^{n} \hat{\tau}_i \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2}{e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}} \boldsymbol{z}_i\,, \tag{25}$$

Further, the second derivatives of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be obtained as:

$$\frac{\partial^2 Q_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \hat{\tau}_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}} \tag{26}$$

$$\frac{\partial^2 Q_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} = -2\sum_{i=1}^{n} \hat{\tau}_i \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})}{e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}} \boldsymbol{x}_i \boldsymbol{z}_i^T\,, \tag{27}$$

$$\frac{\partial^2 Q_c}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} = -2\sum_{i=1}^{n} \hat{\tau}_i \frac{(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2}{e^{2\boldsymbol{z}_i{}^T\boldsymbol{\gamma}}} \boldsymbol{z}_i \boldsymbol{z}_i^T. \tag{28}$$

Then, the Fisher information matrix is

$$H(\boldsymbol{\theta}) = \frac{\partial^2 Q_c}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} \dfrac{\partial^2 Q_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \dfrac{\partial^2 Q_c}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}^T} \\ \dfrac{\partial^2 Q_c}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}^T} & \dfrac{\partial^2 Q_c}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} \end{bmatrix}.$$

4. Repeat these E and M steps until convergence is met.