Normal-Compound Gamma priors with Count Data

Ahmed Alhamzawi[a]*, Gorgees Shaheed Mohammad[b]

[a]*University of AL-Qadisiyah, College of Science, Department of Mathematics, Iraq.Email: ahmed.alhamzawi@qu.edu.iq*

[b]*University of AL-Qadisiyah, College of Education, Department of Mathematics, Iraq.Email: gorgees.alsalamy@qu.edu.iq*

A R T I C L E   I N F O

A B S T R A C T

Count data models have become very common in several disciplines in recent years. Since these types of models can often be studied incorrectly using OLS methods, several solutions have been proposed to address this problem. One of these methods the normal-scale mixture method with different types of priors of the scale parameter. The importance of this method is to solve the issue of the bias-variance tradeoff by adding a local scale parameter to reduce the variance at the origin and reduce the bias at the tails. In this paper, a compound-gamma prior is placed for the scale parameter and the relevant Gibbs sampler is solved for posterior inference. The comparison of the performance of the proposed model with some other existing methods using both very sparse and low sparsity simulated data shows that the proposed model performs very well.

MSC..

1.   Introduction

Consider the linear model

$$y = X\beta + \epsilon, \tag{1}$$

where $y = (y_1, \cdots, y_n)^T, X = (x_1, \cdots, x_p), \beta = (\beta_1, \cdots, \beta_p)^T$ and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$,defined by $\epsilon_i \sim N(0, \sigma^2)$, are the vector of observations, the design matrix of covariates, the vector of unknown regression coefficient and the vector of errors, respectively. In this paper we assume that this model is concerned with count data. The study of Bayesian regression for count data has become an important area in this subject. Several studies have been proposed such as the study of the crash frequency and its influencing factors [17], the study of university credits and it is relationship to pre-enrolment assessment tests [7], the Malaysian motor insurance claim [5], socioeconomic factors and number of tuberculosis [16] and the application of count data to environmental epidemiology [11]. These varies examples show the importance and the need to develop…

∗Corresponding author

Email addresses: ahmed.alhamzawi@qu.edu.iq

The development of count data models in the framework of Bayesian regression was develop in several stages beginning in the 1970s to the modern day [14,6,13]. The most notable of these stages is the work by [9] involving the use of a Poisson process to analysis regression models. Some of modern advances in this arena have involved the introduction of quantiles through conditional quantile functions [10] which requires some particular assumptions about the model while allowing the researcher to analyze the effect of the covariates on each quantile of the distribution [12]. Since the dependent variable y defined by (1) is a generated using a Poisson countable process

$$y_i \sim Poisson(\mu_i)$$

$$\mu_i \sim exp\, exp\, \{x_i^T \beta\}$$

(2)

With only nonnegative integers then it is necessary to convert it to a continuous variable. In order to do this we use the jittering process presented in [12] by adding a uniform variable u_i to our dependent variable y_i then taking the log to produce the desired continuous variable

$$y_i^* = ln\, ln\, \{y_i + u_i\}$$

$$u_i \sim unif(0,1)$$

(3)

In this paper, we will analyze the Bayesian regression framework in the present of count data with a normal scale-mixture combined with a normal-compound gamma prior of the form

$$\beta_j | \sigma^2, z_1 \sim N(0, \sigma^2 z_1)$$

$$z_1 \sim CG(c_1, \dots, c_N, \phi)$$

$$\sigma^2 \sim IG(c_0, d_0)$$

(4)

where $CG(c_1, \dots, c_N, \phi)$ is a compound-gamma distribution of order N and $IG(c_0, d_0)$ is the inverse gamma distribution with shape parameter c_0 and scale parameter d_0.

The paper will be structured as follows: in section 2 the compound gamma prior will be introduced, in section 3 our sampler with be derived, in sections 4 and 5 we will use simulated and real data, respectively, to compare the accuracy of our model with other models.

1.   The normal-compound gamma prior

The compound gamma prior can be written as

$$\pi(x) = \int_0^\infty \dots \int_0^\infty G(z_1, c_1, z_2)G(z_2, c_2, z_3) \dots G(z_N, c_N, z_{N+1})dz_2 \dots dz_N$$

(5)

Or alternatively [1]

$$\pi(x) = \int_0^\infty \dots \int_0^\infty \left[\prod_{i=1}^N \frac{z_{i+1}^{c_i}}{\Gamma(c_i)} z_i^{c_i-1}\{-z_i z_{i+1}\} \right] dz_2 \dots dz_N$$

(6)

where $z_1 = x$ and $z_{N+1} = \phi$ is a constant. A more efficient way is to write this a product of multiple scale mixtures [1] with

$$z_1 \sim G(c_1, 1), z_2 \sim IG(c_2, 1), \dots, z_N \sim \{G(c_N, \phi)\, odd\ N\ IG(c_N, \phi)\, even\ N$$

(7)

where $G(a, b)$ is the gamma distribution with shape parameter a and inverse scale (rate) parameter b.

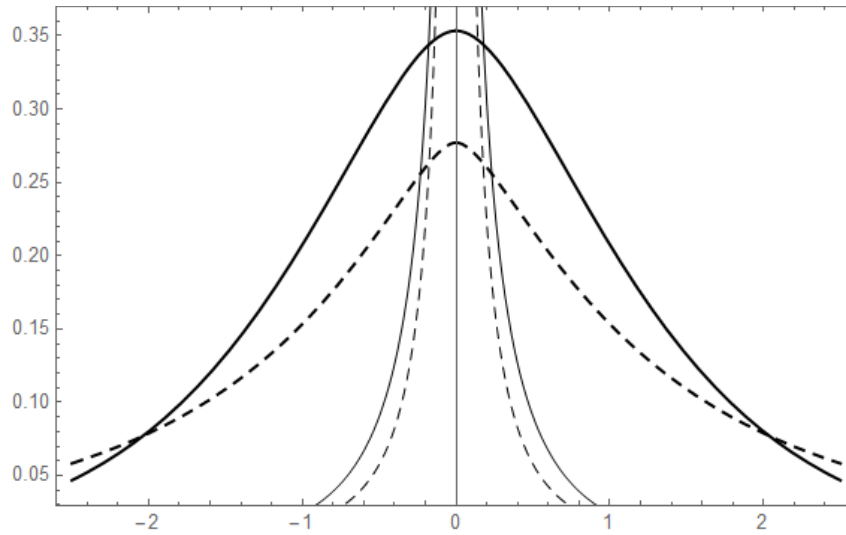**Proof**. The proof of this equivalence is giving in [1].

Figure 1. Plot of the NCG prior with the solid, dashed represent $N = 2$ and $N = 4$, respectively, while the thick and thin lines represent larger and smaller values of c_1, respectively.

The properties of this prior were extensively studied in [1]. Most interestingly, it was shown that this prior is a generalization of various popular models (See [2,15,3] for $N = 2$ and [4] for $N = 4$). In particular, it was shown that this model works for data with different degrees of sparsity. This property is farther demonstrated in Figure 1. The above prior has been investigated in several scenarios and types of data. In this paper, it is our aim to derive the analyze the posterior inference in the area of count data. We notice that for small values of c_1, both in the cases of compounding two and four gamma distributions for there is singularity at zero and distribution mass is concentrated near zero. Thus, this shows how the model works for sparse data. This is shown by the thin lines in the graph. On the other hand, the thick lines show the prior for the non-sparse framework where more the distribution has at both tails of the distribution. In [2], it is shown that for $N = 2$ the EM algorithm can accurate sample the true sparsity (or density) of our data to get more insight to the value of our hyperparameters.

2. Posterior Inference

The full conditionals for $N = 2$ were calculated by [2] to obtain

$$\beta|rest \sim N(\mu_\beta, \Sigma^{-1}\sigma^2) \quad z_1|rest \sim GIG\left(\frac{\beta^T\beta}{\sigma^2 z_2}, 2, c_1 - \frac{1}{2}\right) \quad z_2|rest \sim IG\left(c_2 + \frac{1}{2}, \frac{\beta^T\beta}{2\sigma^2 z_1} + 1\right) \quad \sigma^2|rest$$
$$\sim IG\left(\frac{n + p + 2c_0}{2}, \frac{(y - X\beta)^T(y - X\beta) + \beta^T Z^{-1}\beta + 2d_0}{2}\right)$$

(8)

More generally for $N \geq 2$, the full conditionals are given by [1]

$$\beta|rest \sim N(\mu_\beta, \Sigma^{-1}\sigma^2) \quad z_{odd}|rest \sim GIG\left(\frac{\beta^T Z_{-k}^{-1}\beta}{\sigma^2}, 2, c_k - \frac{1}{2}\right) \quad z_{even}|rest \sim IG\left(c_k + \frac{1}{2}, \frac{\beta^T Z^{-1}\beta}{2\sigma^2} + 1\right) \quad \sigma^2|rest$$
$$\sim IG\left(\frac{n + p + 2c_0}{2}, \frac{(y - X\beta)^T(y - X\beta) + \beta^T Z^{-1}\beta + 2d_0}{2}\right)$$

(9)

where $\mu_\beta = \xi_2^{-2}\Sigma^{-1}X^T V^{-1}(y - \xi_1 v), V = diag(v_1, \ldots, v_n),$ $Z = daig(\prod_{i=1}^N z_{k1}, \ldots, \prod_{i=1}^N z_{kp}), Z_{-k} = daig(\prod_{i=1, i\neq k}^N z_{k1}, \ldots, \prod_{i=1, i\neq k}^N z_{kp})$ and $\Sigma = \xi_2^{-2}X^T V^{-1}X + Z^{-1}$. Therefore, we have the normal distribution $N(\mu_\beta, \Sigma^{-1}\sigma^2)$. For updating the hyperparameters, again we follow [1] using

$$Q(\xi, \xi^{old}) = \sum_{k=1}^{N} \sum_{i=1}^{p} (-1)^{k+1} c_k E_{c_k^{old}}[(z_{ki})|y^*] + c_N(\phi) - \sum_{k=1}^{N} (\Gamma(c_k)) \quad (10)$$
$$+ C$$

where $\xi^{old} = (c_1^{old}, \ldots, c_N^{old})$ and C all the terms not containing $c_1, c_2, \ldots, c_N$. Then, we have

$$\Gamma'(c_k) = \sum_{i=1}^{p} (-1)^{k+1} E_{c_k^{old}}[(z_{ki})|y^*] + c_N(\phi)I(k = N) \quad (11)$$

We will set $c_0 = d_0 = 10^{-5}$ in the prior of σ^2 for estimating the hyperparameters c_k from the EM algorithm. The EM algorithm for finding the values of c_k will be used every few iterations of the MCMC algorithm. We will implement the self-adaptive normal compound model Monte Carlo EM (MCEM) algorithm.

3.   Simulation Studies

To demonstrate the advantages of the proposed model, we will use simulated data to analyze the predicative ability of our prior and compare it to other published models. Specifically, the comparison will be applied with the Beta Prime model ($N = 2$) proposed by [3], our prior $NCG10$ with ($N = 10$), the Bayesian Lasso, the Bayesian adaptive Lasso (aLasso) and the elastic net (Enet). The simulated data will be compared with the mean squared error (MSE), the false positive rate (FPR) and the false negative rate (FNR).

Simulation 1

In our first simulation, we will study the simulated data generated with a very sparse model by setting $\beta = (7, 0, 0, 0, 0, 0, 0, 0)$. From the results presented in Table 1 that averaged from $n = 100$ repeated simulations with 15000 iterations each, we can see that the results show that our prior produces better than other presented methods. Furthermore, we can see that the compound gamma prior gives the smallest MSE compared to all the other models presented. The hyperparameters are updated each 100 iterations using equation (11). Additionally, our model preforms very well in terms FPRs and FNRs which are necessary for selecting the best model for variable selection. From the trace plots and histograms in Figure **2**, we notice how well our prior converges compared to the stationary distribution.

Table 1. Results for Simulation 1.

	MSE (sd)	FPR (sd)	FNR (sd)
NCG2	0.0056 (0.0050)	0.0000 (0.0000)	0.0000 (0.0000)
NCG10	**0.0036 (0.0047)**	0.0000 (0.0000)	0.0000 (0.0000)
Lasso	0.0203 (0.0177)	0.0000 (0.1000)	0.0000 (0.2000)
aLasso	0.0063 (0.0064)	0.0000 (0.0000)	0.0000 (0.4000)
Enet	0.0261 (0.0243)	0.0000 (0.3000)	0.0000 (0.0000)

Simulation 2

To study our model farther and get a deeper view of its behavior with different types of data will a similar size of covariates as the simulation above but we will decrease the sparsity of the model by setting $\beta = (4.5, 2, 0, 3, 0, 0, 0, 8)$ with $n = 100$ repeated simulations with 10000 iterations each. The results are showing in Table 2. We see that the results farther prove what Figure **1** shows, namely, that as the sparsity increase the model performs better higher values of N and conversely, models with smaller values of N are better candidates for data with less sparsity. Similar

to simulation 1, to the trace plots and histograms in Figure **3** show that our mode converges better than the stationary distribution. Similarly, the hyperparameters are updated each 100 iterations.

Table 2. Results for Simulation 2.

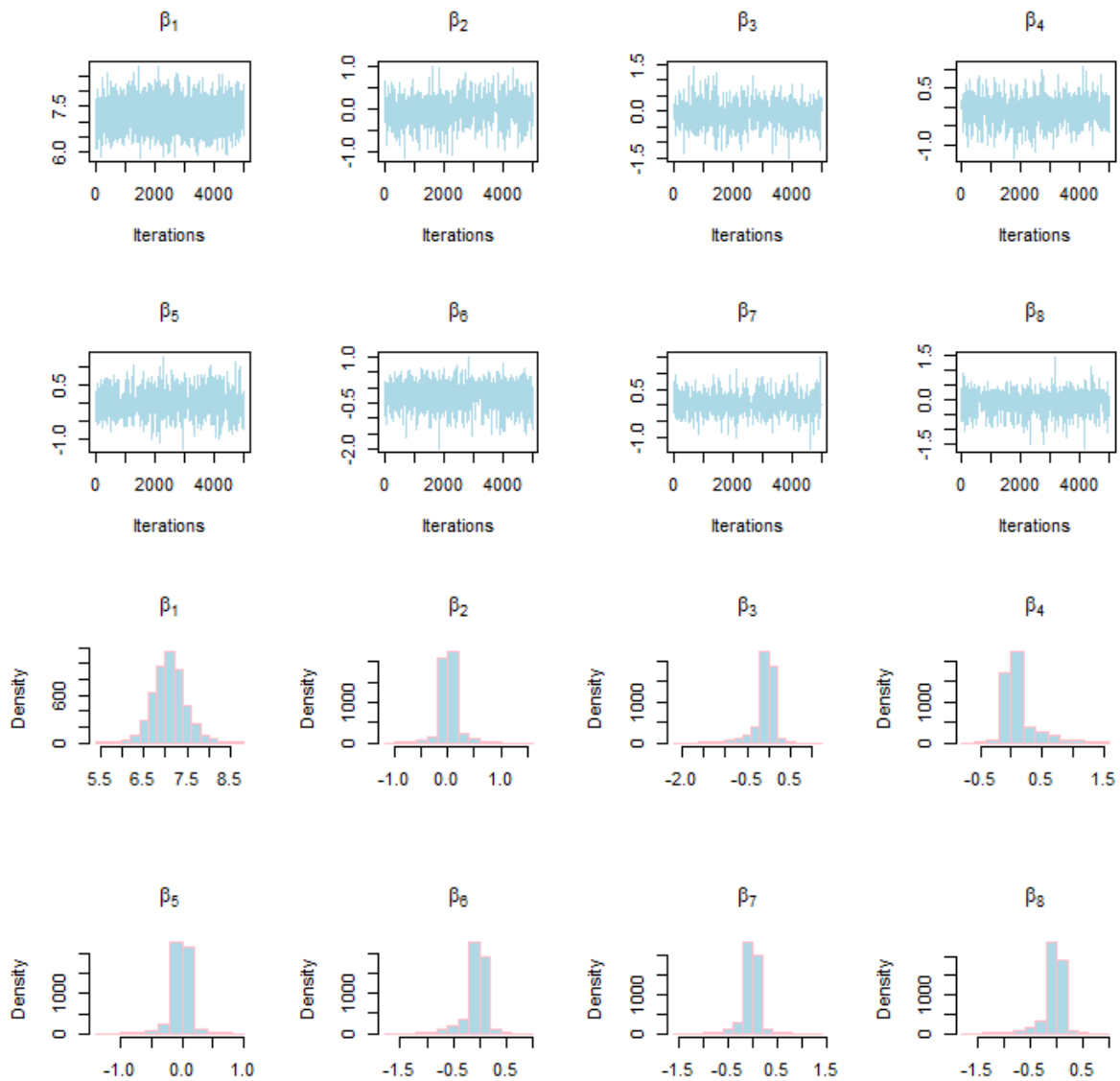|	MSE (sd)	FPR (sd)	FNR (sd)
NCG2	**0.0206 (0.0122)**	0.2100 (0.2638)	0.4100 (0.4312)
NCG10	0.0304 (0.0134)	0.2300 (0.2723)	0.4600 (0.5037)
Lasso	0.0662 (0.0156)	0.2900 (0.4793)	0.4900 (0.5126)
aLasso	0.0401 (0.0160)	0.2100 (0.4320)	0.4300 (0.5252)
Enet	0.0699 (0.0208)	0.2600 (0.5912)	0.4200 (0.5523)

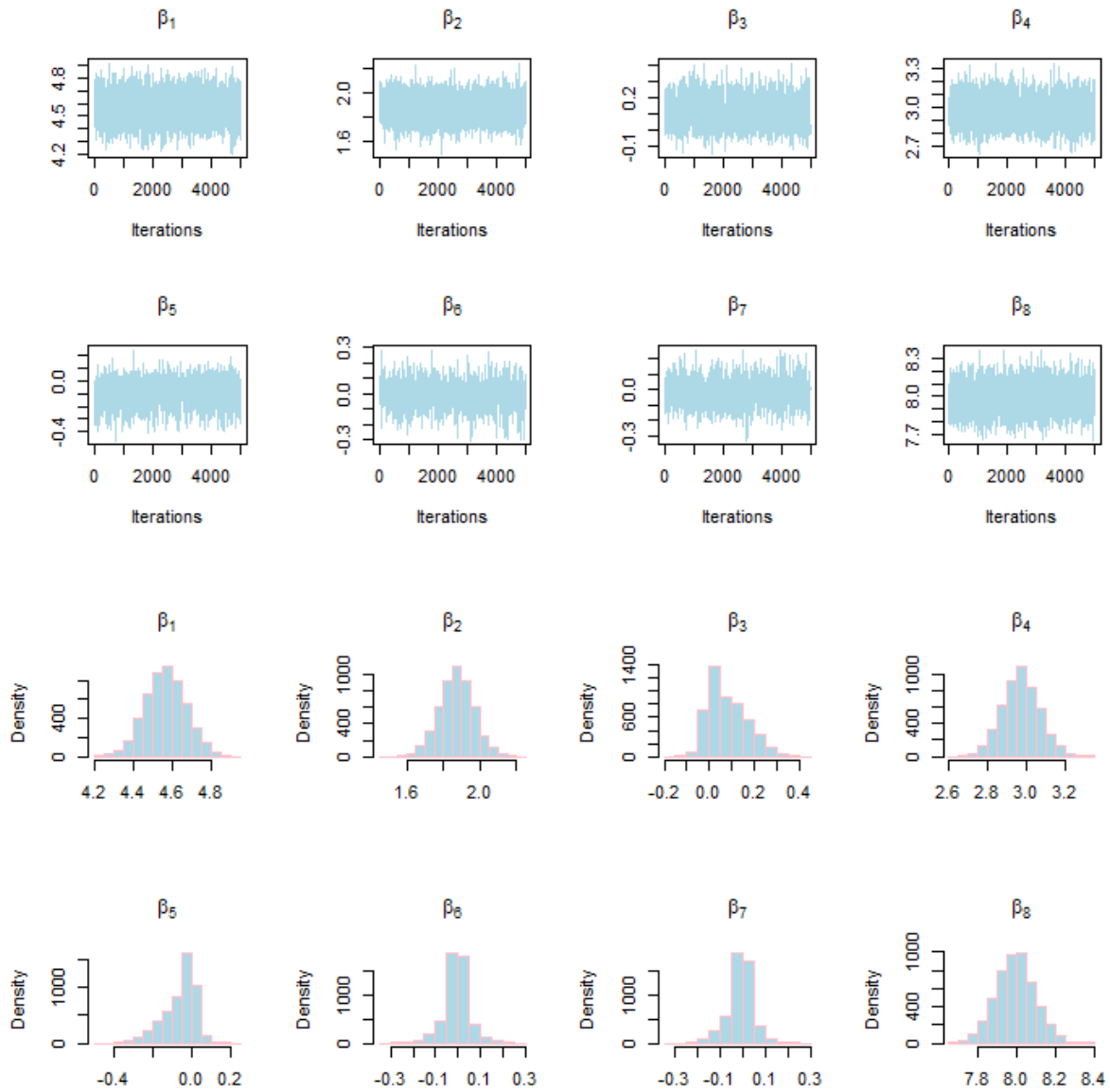**Figure 2.** Trace plots (top) and histograms (bottom) for simulation 1.

Figure 3. Trace plots (top) and histograms (bottom) for simulation 2.

4.   Conclusion

We have shown that our proposed method for count data using normal-compound gamma prior for the scale mixture in the framework of normal-scale mixture perform very well compared to other existing models such as the Beta Prime prior (NCG2) [3], The Bayesian lasso, the Bayesian adaptive lasso and the Bayesian elastic net. We aim to study this method further in the future with different types of data such as censored and quantile data.

References

[1]   Alhamzawi, A., & Mohammad, G. S. (2022). Statistical inference with normal-compound gamma priors in regression models. arXiv preprint arXiv:2212.13144.

[2]   Armagan, A., M. Clyde, and D. Dunson (2011). Generalized beta mixtures of gaussians. Advances in neural information processing systems 24.

[3]   Bai, R. and M. Ghosh (2018). On the beta prime prior for scale parameters in high dimensional bayesian regression models. arXiv preprint arXiv:1807.06539.

[4]   Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. Biometrika 97 (2), 465–480.

[5]   Fuzi, M. F. M., Jemain, A. A., & Ismail, N. (2016). Bayesian quantile regression model for claim count data. Insurance: Mathematics and Economics, 66, 124-137.

[6]   Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica: Journal of the Econometric Society, 701-720.

[7]   Grilli, L., Rampichini, C., & Varriale, R. (2016). Statistical modelling of gained university credits to evaluate the role of pre-enrolment assessment tests: An approach based on quantile regression for counts. Statistical Modelling, 16(1), 47-66.

[8]   Jantre, S. (2023). Bayesian quantile regression for longitudinal count data. Journal of Statistical Computation and Simulation, 93(1), 103-127.

[9]   Jorgenson, D. W. (1961). Multiple regression analysis of a Poisson process. Journal of the American Statistical Association, 56(294), 235-245.

[10]   Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.

[11]   Lee, D., & Neocleous, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. Journal of the Royal Statistical Society: Series C (Applied Statistics), 59(5), 905-920.

[12]   Machado, J. A. F., & Silva, J. S. (2005). Quantiles for counts. Journal of the American Statistical Association, 100(472), 1226-1237.

[13]   McCullagh, P., and Nelder, J. A. (1989), Generalized Linear Models (2nd ed.), London: Chapman & Hall.

[14]   Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), 370-384.

[15]   P´erez, M.-E., L. R. Pericchi, and I. C. Ram´ırez (2017). The scaled beta2 distribution as a robust prior for scales. Bayesian Analysis 12 (3), 615–637.

[16]   Sarvi, F., Momenian, S., Khodadost, M., Pahlavanzadeh, B., Nasehi, M., & Sekhavati, E. (2016). The examination of relationship between socioeconomic factors and number of tuberculosis using quantile regression model for count data in Iran 2010-2011. Medical Journal of the Islamic Republic of Iran, 30, 399.

[17]   Wu, H., Gao, L., & Zhang, Z. (2014). Analysis of crash data using quantile regression for counts. Journal of Transportation Engineering, 140(4), 04013025.