

*The statistical control chart is considered one of the superlative tools in quality control. Currently, control charts are being widely used in various areas, one of them being manufacturing processes. They are essential instruments that can impart crucial insights to quality controllers for maintaining productivity. The quality of a product or process can be characterized by a relationship between two or more variables, which is typically referred to as a profile. Also, public health surveillance is considered another important area that widely used control charts. In this regard, they are very useful and reliable tools for detecting outbreaks of infectious diseases. On the other hand, the gamma regression model (GRM) is a popular model considered in medical and other fields. It is applied when the response variable is continuous and positively skewed and well fitted to the gamma distribution. This paper presents a scheme for monitoring the profile. Based upon the generalized linear model (GLM) in the case of two link functions: identity and log link function. Exponentially weighted moving average control charts (EWMA) are proposed using deviance residuals and Pearson residuals for detecting any disturbance in the control variable of the gamma regression model. A detailed simulation study is designed to scrutinize and evaluate the performance of the control charts in phase I analysis and in phase II under parametric maximum likelihood estimation (MLE) using the average run length (ARL) measure. It turns out that using deviance residuals under the identity link function seems more suitable than Pearson residuals. Also, with increasing the sample size, the percentages of out-of-control (OC) samples increased which is theoretically acceptable*

**Keywords:** ARL, Control Charts, EWMA, GLM, Gamma Regression, Residuals

# PROFILE MONITORING OF RESIDUALS CONTROL CHARTS UNDER GAMMA REGRESSION MODEL

**Salah M. Mohamed**

Doctor of Applied Statistics, Professor\*

**Engy S. Abozaid**

Master of Statistics, Lecturer\*

**Shereen H. Abdel Latif**

Corresponding author

PhD, Assistant Professor\*

E-mail: shereen\_hamdy\_cu.edu.eg

\*Department of Applied Statistics  
and Econometrics

Cairo University

Dr. Ahmed Zewail str., 5, Orman,

Giza, Egypt, 12613

Received date 21.09.2022

Accepted date 08.12.2022

Published date 30.12.2022

**How to Cite:** Mohamed, S. M., Abozaid, E. S., Abdel Latif, S. H. (2022). Profile monitoring of residuals control charts under gamma regression model. *Eastern-European Journal of Enterprise Technologies*, 6 (4 (120)), 23–31. doi: <https://doi.org/10.15587/1729-4061.2022.264904>

## 1. Introduction

One of the most powerful tools in statistical process control (SPC) is the statistical control charts, the idea of graphical summary of information for monitoring and controlling of the production process [1] presented the term of «profile» as refer to the relationship between the response and one or more explanatory variable(s). [2] proposed the fundamental objective of profile monitoring is to check the stability of the functional relationship between the response and the explanatory variable(s) over time. Many studies have been done by researchers for monitoring different types of profiles such as: gamma profiles, Poisson profile and so on. In practice, control chart analyses are carried out either in phase I or phase II. In phase I, it is often interested in checking the stability of a process and identifying possible out-of-control samples. But, in phase II it is often interested in detecting an out-of-control condition fast. Average run length and probability of signal are the common criteria to evaluate performance of the control schemes in phases I and II, respectively. [3] provided a unified framework for phase I analysis control of generalized linear profiles. Besides the generalized linear models (GLMs), other types of models have also been used to represent profiles, such as simple linear regression, [4] and multiple regression, [5].

For detecting any disturbance in the control variable of the gamma regression model exponentially weighted moving average control charts (EWMA) are proposed. This methodology, EWMA, was first introduced by [6] who evaluated its properties and showed that the EWMA is an effective alternative to the traditional Shewhart control chart when small shifts in the process parameters are monitored. [7] also reached the same results. Various modification and supplemental criteria have also suggested that control charts based on moving average is also an effective chart in detecting small process change [8]. [9] stated that the Shewhart charts are slow in detecting small shifts as they use only the information contained in the last sampling point. The EWMA and cumulative sum (CUSUM) charts are time weighted control charts that were suggested as superior alternatives to the Shewhart chart when the detection of small shifts was desired. [10] studied the monitoring of EWMA control charts to obtain accurate average run length (ARL) values of  $(\bar{x}, s^2)$  and EWMA control charts. [11] stated that Shewhart control charts are used to detect large shifts in a process, whereas the CUSUM chart and the EWMA chart are more efficient to detect shifts from small to moderate in a process. [12] defined EWMA chart as a time weighted and a powerful tool for detecting small shifts in a parameter of a process more rapidly than Shewhart chart with an equal sample size.

Therefore, in practice, gamma distribution can be used with great flexibility in the analysis of positive random variables. Thus, gamma regression models (GRM) are applied in a wide range of empirical applications, such as in the process of rate setting in the framework of heterogeneous insurance portfolios, which is the most important function of insurers, and in hospital admissions for rare diseases, [13, 14]. [15] proposed the use of GLMs – based control charts for monitoring gamma distributed response variables. The procedure was only based on the deviance residual, which is a likelihood ratio statistic for detecting a mean shift when the shape parameter is assumed to be unchanged, and the input and output variables were related in a certain manner.

The statistical control chart is considered as they are essential instruments that can impart crucial insights to quality controllers for maintaining productivity. Also, public health surveillance is considered another important area that widely used control charts. In this regard, they are very useful and reliable tools for detecting outbreaks of infectious diseases. Therefore, research on monitoring the profile of different types of residuals in phase I and phase II analysis for very important regression model such as GRM which is considered one of GLM is relevant.

---

## 2. Literary analysis and statement of the problem

---

The importance of the study comes in relation to the subject of SPC research, charts are essential tools that are used to monitor process quality. Based on these charts, operators may immediately spot changes in quality characteristics and pursue high-quality production or service. Traditional control chart research typically assumes that quality characteristics follow a normal distribution, but, non-normal response outcomes can occasionally occur in processes, such as discrete count data. Traditional control charts cannot process non-normal data due to the distributional assumption constraint; thus statistical profile monitoring of non-normal response outcomes is desirable.

For this purpose, it is aimed to analyse the usage of residual control charts in conjunction with forecasting models to evaluate the features of production processes. The primary goal is to assess the effectiveness of exponentially weighted moving average (EWMA) and control charts for individual observations (CCIO) charts when applied to residuals of models to identify outliers in auto correlated processes.

[16] designed and studied the performance of the EWMA control charts that monitor the rate of road crashes. They observed that EWMA control chart scheme is sensitive in detecting small shift in the process and it detects shift quickly for this type of dataset. And [17] proposed two types of residuals for gamma regression model; for which many link functions can be used, so they chose the identity and log link for their evaluation. As it is known, the residual analysis pursues to specify outliers and/or model misspecification. And it can be based on ordinary residuals, standardized variants, or deviance residuals. i.e., most residuals are based on the differences between the observed responses and the fitted conditional mean. In this regard the authors constructed more reliable goodness of fit measures and measures of explained variation for gamma regression models. [18] and [19] evaluated EWMA control chart based on the residuals for autocorrelated observations. Because they can have a significant effect on the control chart performance.

[20] focused on monitoring a process that is measured by a linear function and proposed to monitor the average of the residuals for the samples with EWMA and R-chart. [21] used EWMA control charts to monitor a drilling process to detect chatter vibration and to secure production with high quality. These control charts used the residuals obtained from an approximated autoregressive model [22] compared the effectiveness of the Shewhart  $\bar{x}$ , EWMA, and geometric moving average (GMA) residual control charts for auto-correlation observations. They showed that the performance of EWMA charts was superior to the Shewhart  $\bar{x}$ , and GMA residual charts for small shifts, however the performance of Shewhart  $\bar{x}$  residual chart was superior to EWMA and GMA residual charts for large shifts.

[23] designed EWMA chart of Pearson residuals of a negative binomial regression. They founded that their proposed charts present better performances over EWMA charts for deviance residuals, with a remarkable advantage of the Pearson residuals, which are much easier to interpret and calculate. [24] applied three well-known count data models; Poisson, negative binomial, and Conway-Maxwell-Poisson to identify the best fit model for the number of crashes. Conway-Maxwell-Poisson was identified as the best fit model, GLM-based EWMA and CUSUM control charts were proposed using the randomized quantile residuals and deviance residuals. Their simulation study was designed for predictive performance evaluation of the proposed control charts with Shewhart charts. Their study results showed that the EWMA type control charts have better detection ability compared with the CUSUM type and Shewhart control charts under small and/or moderate shift sizes. [25] proposed GLM – based control charts for inverse Gaussian response variable. Deviance and Pearson residuals of the IG regression were considered as plotting statistics. An example related to the yarn manufacturing industry and a simulation study was designed and the performance of the proposed methods was compared with existing counterparts in terms of the run length properties. Moreover, run-rules were implemented to gain the efficiency of the Shewhart type GLM-based control charts under small-to-moderate shifts.

Most of the reviewed studies assessed EWMA performance of residuals for types of variables/models that differ from the gamma variable/model such as count variable or time series model [16, 18–25].

So, it is proposed to supplement the work of [17] via designing EWMA charts of residuals of a gamma regression model. Also, it is proposed to design EWMA chart of Pearson residuals such as [23] and [25] because of the aforementioned advantages, to evaluate its performance in relation to EWMA charts for deviance residuals.

All this suggests that it is advisable when conducting a study on the control variable of two different gamma regression models to depend on exponentially weighted moving average control charts (EWMA) for some different kinds of residuals such as deviance and Pearson residuals to detect any disturbance in that variable since EWMA is superior to any existing counterparts, especially when it is desired to detect small and moderate shifts using the average run length (ARL) or RARL measures.

---

## 3. The aim and objectives of the study

---

The aim of this study is to determine which of the proposed charts is an efficient chart than its existing competitive

control charts for detecting out-of-control process quickly. The gamma distribution has been used to model, for example, the time between events, the size of insurance claims and rainfalls, and failure times of machine parts.

To achieve this aim, the following objectives are accomplished:

- to investigate the performance of the two different types of residuals via EWMA charts by using RARL measure for first gamma model with the identity link function;
- to investigate the performance of the two different types of residuals via EWMA charts by using RARL measure for second gamma model with the log link function;
- to compare the performance of the two different gamma models with three different sample sizes: four, five, and ten via EWMA charts by using RARL measure;
- to compare the performance of the two different gamma models with three different numbers of samples: twenty-five, thirty, and thirty-five via EWMA charts by using RARL measure.

---

#### 4. Materials and methods of the study

---

##### 4. 1. Object and hypothesis of the study

Simulation modeling was carried out to compare between performances of two residuals control charts: deviance residuals control charts and Pearson residuals control charts for two gamma regression models. To evaluate their performance under parametric estimation method maximum likelihood estimation (MLE) by using two types of control charts designs: x-bar/s charts in phase I analysis and the EWMA control charts in phase II analysis using the average run length (ARL) measure. According to the simulation study data, it is generated by fitting two different gamma regression model using the maximum likelihood estimation (MLE) as a parametric regression estimation method in case of two different link functions; identity and log link functions. Then, extract two types of residuals: Pearson residuals and deviance residuals. After that, the exponentially weighted moving average control chart is design to monitor residual types by using x-bar/s chart in phase I analysis and using EWMA chart in phase II analysis. Then the ARL and RARL are computed for EWMA control charts.

##### 4. 2. Generalized linear models

Generalized linear model term was introduced by [26] who extend the scoring method to maximum likelihood estimation in exponential families. [27] proposed GLMs model as a class of regression models appropriate for application in the case of non-normally distributed response variables. These models based on probability distributions with an unknown location parameter  $\varphi$ , if it belongs to the exponential family. The probability density function of the exponential family is most seen, as in (1):

$$f(y; \varnothing, \varphi) = \exp[(y(\varnothing) - b(\varnothing)) / a(\varphi) + c(y, \varphi)]. \quad (1)$$

For some specific functions  $a(\cdot)$ ,  $b(\cdot)$ , and  $c(\cdot)$ . If  $\varphi$  is known, then the family is termed the linear exponential family and  $\varphi$  is the natural or canonical parameter. From (1) there are  $E(y) = b'(\varphi)$  where  $b'(\varphi) = db(\varphi)/d\varphi$ , and  $var(y) = a(\varphi)b''(\varphi)$ .

GLMs models are structured around three components.

a) the systematic component: this specifies the explanatory variables as a set of linear predictors. The linear pre-

dictor, denoted by  $\eta$  is the non-random component which given by:

$$\eta = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik},$$

or can be written in matrix notation  $\eta = X\beta$ .

b) the random component: this identifies the probability distribution of the response variable in which  $y$  belongs to the exponential family as independent random variables having the same form of distribution. Apart from the normal, other distributions such as binomial, Poisson, and Gamma can be handled;

c) the link function: this specifies the relationship between the systematic component and value expected of random component. It says how the expected value of the response relates to the linear predictor of explanatory variables. The link function can be expressed as:

$$\mu = \eta = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}, \quad (2)$$

where  $\beta$ 's are unknown coefficients and  $k$ 's are control variables, the application of the GLMs for modelling data process allows for better precision of estimates, nonlinear relationship between the variables and predicting their behaviour. The systematic component that makes up the regression model is the structure of control variables as a linear sum  $\eta$  and the relationship between variables in a GLMs can be expressed by a known function  $g(\cdot)$ , called the link function, which  $g(\cdot)$  denotes a fixed link function between the mean of the response variable and the linear combination of the explanatory variables (the linear predictor  $\eta$ ).

##### 4. 3. Gamma regression models

[28] studied the gamma distribution that can be used for regression models with more flexibility than other models, such as the exponential and Poisson, and proposed GRM assumed that the pendent variable is gamma distributed and that it's mean is related to a set of regressors can be identity, the inverse, or the logarithm function. The model also includes a shape parameter, which may be constant or dependent on a set of regressors through a link function, and they considered GRM in with both mean and shape parameters are allowed to depend on covariates, in which these two parameters follow regression structures, are proposed in [29] under both classic and Bayesian approaches.

The Gamma distribution can be viewed as a generalization of the exponential distribution with mean  $1/\psi$ ,  $\psi > 0$ . An exponential random variable with mean  $1/\psi$  represents the waiting time until the first event to occur, where events are generated by Poisson process with mean  $\psi$ , while the gamma random variable  $y$  represents the waiting time until the  $i^{\text{th}}$  event to occur. The probability density function of  $y$  is given by:

$$f(y|\theta, \psi) = (\psi^\theta y^{\theta-1} e^{-\psi y} / \Gamma(\theta)) I(0, \infty)(y), \quad (3)$$

where  $\theta, \psi > 0$ ,  $\Gamma(\cdot)$  denotes the gamma function, and  $I(\cdot)$  is the indicator function.

##### 4. 4. Gamma Regression Residuals

Most residuals are based on the differences between the observed responses and fitted conditional mean. [30] mentioned that the analysis of residuals plays a major role in predictive model formation. When numerical-valued functions

are fitted to sample data, all the information about lack of fit is contained in the residuals which can be used to provide necessary feedback on the modelling process. These residuals from a fitted model can be plotted to help detect unequal variances, relationships over time. There are several references to residual analysis can be found in [31].

Residuals may be employed as measure for model selection, play a significant role in model diagnostics and assessment of its fitting and used for testing the validity of the assumptions of statistical models. For example, residuals are used to verify homoscedasticity, linearity of effects, normality, and independence of the error. In classical linear models, residuals are usually standardized so that they become scale free and have the same precision, and this makes it more convenient to compare residuals at various locations in the region of experimentation. There are different types of residuals such as classical residuals, Pearson residuals, and deviance residuals.

**4. 4. 1. Pearson Residuals**

Residuals in GLMs were first discussed by [32], though ostensibly concerned with logistic regression models. Pearson residuals are the most used measures of overall fit for GLMs, which can be used to check the model for each observation, [33] Defined Pearson Residuals in GRM as follows:

$$r_i^p = (y_i - \hat{\mu}_i) / (\sqrt{\text{var}(\hat{\mu}_i)}) = (y_i - \hat{\mu}_i) / (\sqrt{\hat{\mu}_i (n_i - \hat{\mu}_i) / n_i}), i = 1, 2, \dots, n, \tag{4}$$

where  $\hat{\mu}_i$  is the fitted mean value and the denominator follows from the fact that  $y_i = n_i \pi_i (1 - \pi_i)$  the fitted variance function of  $y_i$ .

**4. 4. 2. Deviance residuals**

[17] defined alternative residual that was based on the deviance or likelihood ratio, which for GRM is given by:

$$r_i^d = -2 \sum_{i=1}^n \log(y_i / \hat{\mu}_i) - ((y_i - \hat{\mu}_i) / \hat{\mu}_i), \tag{5}$$

where  $\hat{\mu}_i = g^{-1}(x_i' \hat{\beta})$ .

Two statistics used to assess the goodness of fit of the GLMs are the deviance and the Pearson chi squared statistic  $\chi^2$ . Linear models use raw residuals for testing and model diagnostic, whereas the GLMs provide several structures for residuals such as the Pearson, deviance, and likelihood. The two common types of residuals are then the components of these statistics. Other definitions for residuals in GLMs have been proposed in [34].

**5. Results of the study profile monitoring of residual control charts under gamma regression model**

**5. 1. Generating the data**

To conduct the simulation study, the observations  $X_i$ ,  $i=1, 2, 3$  were generated such that, the values of  $X_1$ ,  $X_2$ , and  $X_3$  were generated from a uniform distribution on the intervals (0, 30), (0, 15), and (10, 20), respectively. Gamma regression model was built. This model has two different link functions: identity link function and log link function [17]. Different sample sizes ranges from moderate to large were chosen such as 125, 150, 175, 250, 300, and 350. The following are the steps by more details.

First Step: Mean Structure Function with Identity Link Function:

a) the gamma regression model with mean and shape structures given by formulas (6) and (7), respectively, is considered:

$$\mu_i = x_i' \theta, \tag{6}$$

$$\log(\theta_i) = Z_i' y, \tag{7}$$

where (6) describes the variable of interest  $Y$  and the regressors  $X$  of the mean structure. And (7) describes the regressors  $Z$  for the shape structure, with log-link function.

b) the values  $Y_i$  were generated from a gamma distribution with mean and shape parameters given  $\mu_i = 15 + 2x_{2i} + 3x_{3i}$ , and  $\theta_i = \exp(0.2 + 0.1x_{2i} + 0.3x_{3i})$ .

Second Step: Mean Structure Function with Log Link Function:

a) the gamma regression model with mean and shape structures given by (7), (8), respectively, is considered:

$$\log(\mu_i) = x_i' \theta. \tag{8}$$

b) the values  $y_i$  were generated from a gamma distribution with mean and shape parameters given by:

$$\mu_i = \exp(-5 + 0.2x_{2i} - 0.03x_{3i})$$

and

$$\theta_i = \exp(0.2 + 0.1x_{2i} + 0.3x_{3i}),$$

where  $-5$ ,  $0.2$ , and  $-0.03$  are the value of the  $\theta_0$ ,  $\theta_1$ , and  $\theta_2$  coefficients respectively.

**5. 2. Design of the exponentially weighted moving average control chart (Phase II analysis)**

To construct an EWMA chart, the design parameters of it are  $l$  and  $\lambda$ , the multiple of  $\sigma_{z_i}$  used in the control limits and the weighting or smoothing constant, respectively.

The following are steps to conduct the EWMA chart:

a) compute the values of parameters form historical data through X-bar/S-chart such as the target mean, or process mean ( $\mu_0$ ) and the estimated sigma ( $\hat{\sigma}$ ) by using the S-bar method;

b) calculate the EWMA statistic ( $Z_i$ ) and calculate the control limits ( $UCL, LCL$ ) with the combination of ( $l=2.962, \lambda=0.2$ ) for each sample( $i$ ) [8];

c) plot the statistic ( $Z_i$ ) for each sample, with the control limits in place;

d) declare the process to be in-control if  $LCL_{(i)} \leq Z_i \leq UCL_{(i)}$ ; otherwise, declare the process to be OC.

If the process is declared OC, count the number of subgroups as the run length, i.e., the process remains in-control (IC) before it is declared to be OC. Then the ARL for EWMA control charts is computed.

**5. 3. Exponentially weighted moving average control charts results**

**5. 3. 1. Number of samples equals twenty-five**

The following Figures from Fig. 1–4 showed the EWMA values for the 25 samples for identity link function under deviance residuals (ID) and under Pearson residuals (IP), and for log link function under deviance residuals (LD) and under Pearson residuals (LP) respectively.

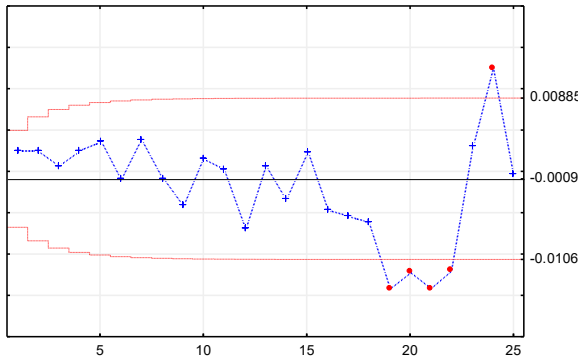


Fig. 1. Exponentially weighted moving average values for 25 samples for identity link function under deviance residuals

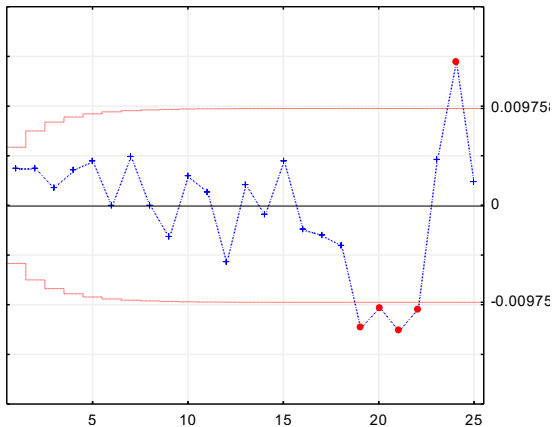


Fig. 2. Exponentially weighted moving average values for 25 samples for identity link function under Pearson residuals

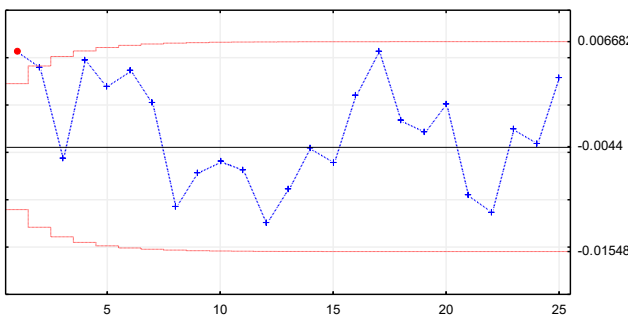


Fig. 3. Exponentially weighted moving average values for 25 samples for log link function under deviance residuals

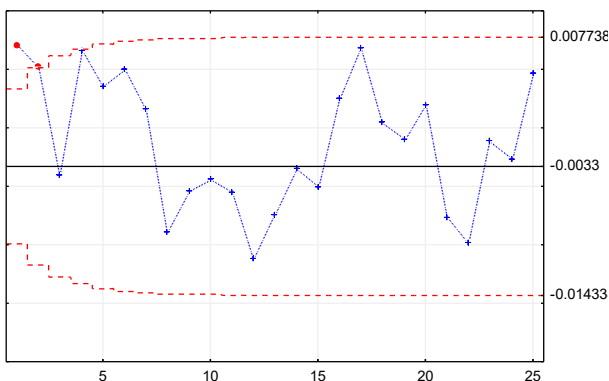


Fig. 4. Exponentially weighted moving average values for 25 samples for log link function under Pearson residuals

Under identity link function, Fig. 1 showed that (16 %) OC lower and (4 %) OC upper and Fig. 2 gave the same result as Fig. 1. But under log link function, Fig. 3 showed that for deviance 4 % of samples became OC (upper) and Fig. 4 showed that for Pearson 8 % of samples became OC (upper).

**5. 3. 2. Number of Samples Equals Thirty**

The following charts from Fig. 5–8 showed the EWMA values for the 30 samples under ID, IP, LD, and LP respectively.

Under identity link function, Fig. 5 showed that (7 %) OC lower and (3 %) OC upper for all samples and Fig. 6 gave the same result as Fig. 5. But under log link function, Fig. 7 showed that for deviance (7 %) of samples became OC lower and (10 %) of samples became OC upper. And Fig. 8 gave the same result as Fig. 7.

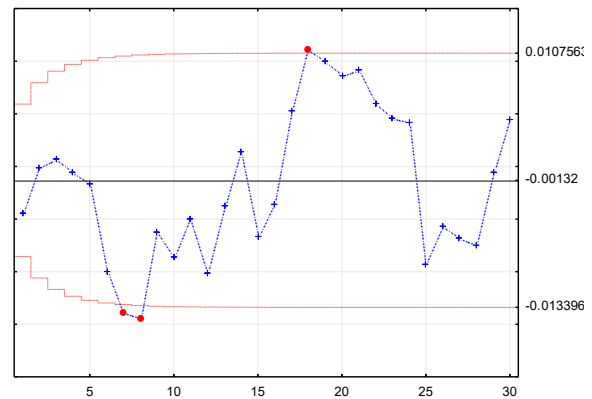


Fig. 5. Exponentially weighted moving average values for 30 samples for identity link function under deviance residuals

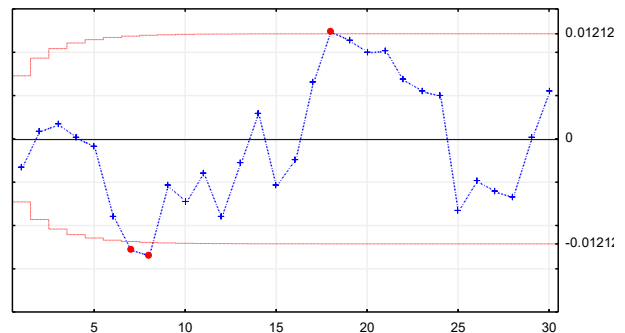


Fig. 6. Exponentially weighted moving average values for 30 samples for identity link function under Pearson residuals

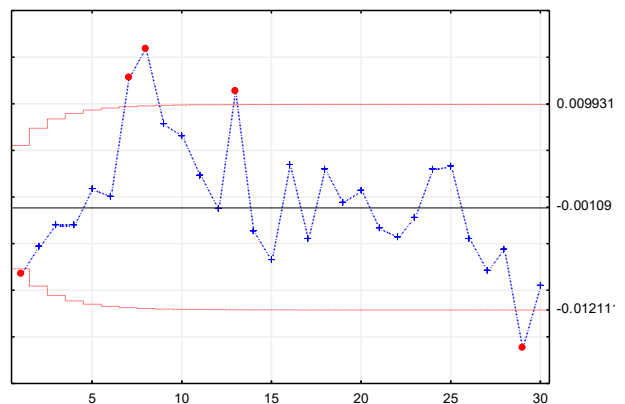


Fig. 7. Exponentially weighted moving average values for 30 samples for log link function under deviance residuals

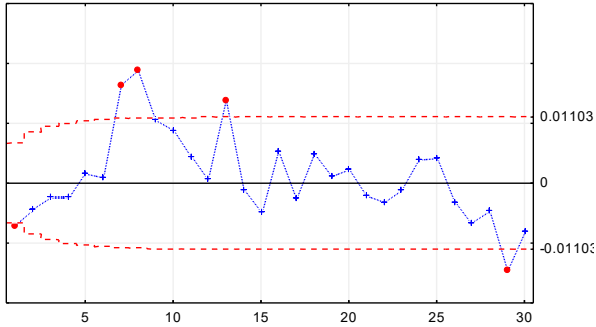


Fig. 8. Exponentially weighted moving average values for 30 samples for log link function under Pearson residuals

**5. 3. 3. Number of samples equals thirty-five**

The following charts from Fig.9–12 showed the EWMA values for the 35 samples under ID, IP, LD, and LP respectively.

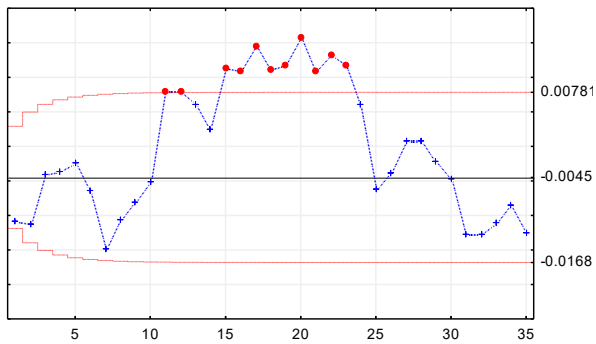


Fig. 9. Exponentially weighted moving average values for 35 samples for identity link function under deviance residuals

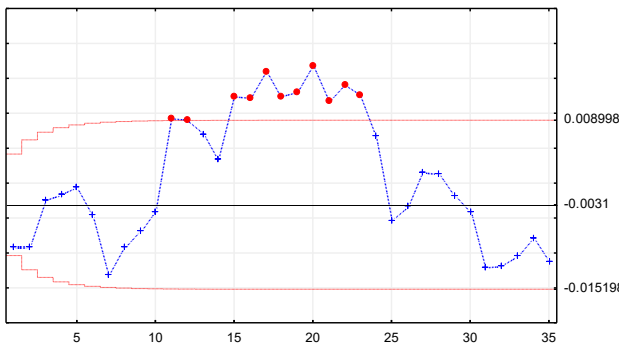


Fig. 10. Exponentially weighted moving average values for 35 samples for identity link function under Pearson residuals

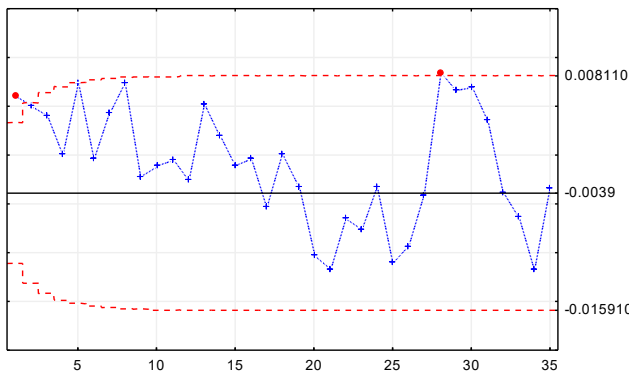


Fig. 11. Exponentially weighted moving average values for 35 samples for log link function under deviance residuals

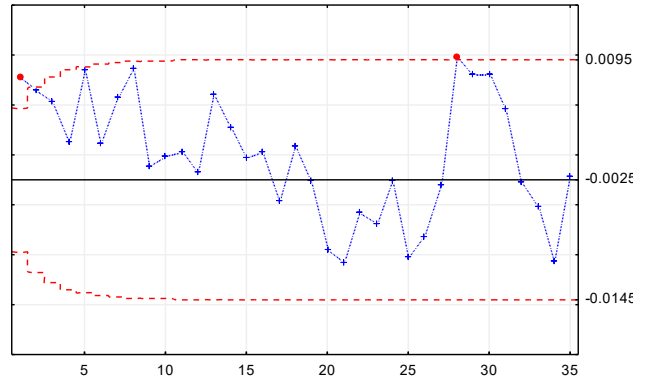


Fig. 12. Exponentially weighted moving average values for 35 samples for log link function under Pearson residuals

Under identity link function, Fig. 9 showed that (31 %) OC (upper) of all samples and Fig. 10 gave the same result as Fig. 9. But under log link function, Fig. 11 showed that 6 % of samples became OC (upper) and Fig. 12 gave the same result as Fig. 11.

**5. 4. Performance of exponentially weighted moving average charts**

The following tables, Tables 1, 2, contain the results of two measures of performance for the charts: ARL and the relative ARL (RARL) for gamma regression models. Three different sample sizes were selected, four, five, and ten.

Table 1

Average run length results

Sample size ( <i>n</i> )	Number of sample ( <i>m</i> )	Identity Link Function		Log Link Function	
		Deviance Residuals	Pearson Residuals	Deviance Residuals	Pearson Residuals
<i>n</i> = 4	25	488.9827	488.967	495.7826	495.7935
	30	495.7225	495.7935	495.7232	495.792
	35	495.7127	495.7929	495.7502	495.7935
<i>n</i> = 5	25	495.7381	495.7935	495.7469	495.793
	30	495.7756	495.7935	495.7451	495.7935
	35	495.7695	495.7931	495.7757	495.7932
<i>n</i> = 10	25	495.7789	495.7935	495.7465	493.7935
	30	495.7526	495.7935	495.7934	494.7935
	35	495.7545	495.7935	495.7654	495.7935

To facilitate the comparison between different methods of estimating the ARL, let's use the following formula for RARL. RARL= Proposed estimate value/Deviance residuals under identity link function.

The following Table 2 showed the relative performance of ARL for all ID, IP, LD, and LP values relative to ID value.

As depicted by Table 2, and by comparing the RARL values for deviance and Pearson residuals under the two kinds of the link functions. It can be observed that, when using identity link function, the ARL of deviance residual is slightly less than the Pearson residual with various choices of sample sizes; 4 or 5 or 10. Also, it is observed that the largest (worst) value of ARL was for Pearson residuals, specifically with *n*=4 or 5 and for both kinds of link function. And the smallest (best) values of ARL was for Pearson resi-

duals (6 %) from all ARL values under identity link function and approximately (11 %) from all ARL values under log link function. But most of the smallest ARL values have been executed under deviance residual (94 %) of ARL values under identity function and (89 %) of ARL values under log link function.

Table 2

Relative average run length results

Sample size ( <i>n</i> )	Number of sample ( <i>m</i> )	Identity link function		Log link function	
		Deviance residuals	Pearson residuals	Deviance residuals	Pearson residuals
<i>n</i> =4	25	1	0.999968	1.013906	1.013929
	30	1	1.000143	1.000001	1.00014
	35	1	1.000162	1.000076	1.000163
<i>n</i> =5	25	1	1.000112	1.000018	1.000111
	30	1	1.000036	0.999938	1.000036
	35	1	1.000048	1.000013	1.000048
<i>n</i> =10	25	1	1.000029	0.999935	0.995995
	30	1	1.000083	1.000082	0.998065
	35	1	1.000079	1.000022	1.000079

For more clarification, the relative average run length values have been shown graphically by Fig. 13–16.

From Fig. 13, it can be concluded that, at *m*=25 there were the worst values for EWMA of log link function for both deviance and Pearson residuals. And from Fig. 14, it can be concluded that, at *m*=30 there were the best values of EWMA for log link functions and for deviance residuals.

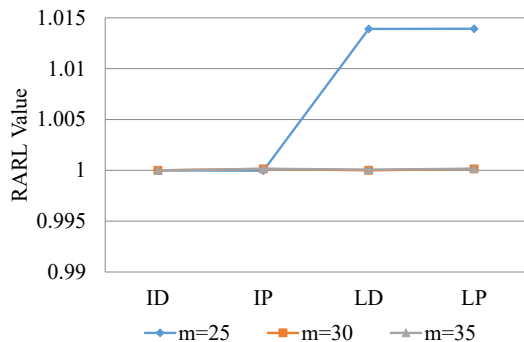


Fig. 13. Relative average run length of exponentially weighted moving average values at *n*=4 of identity and log link functions for both deviance and Pearson residuals

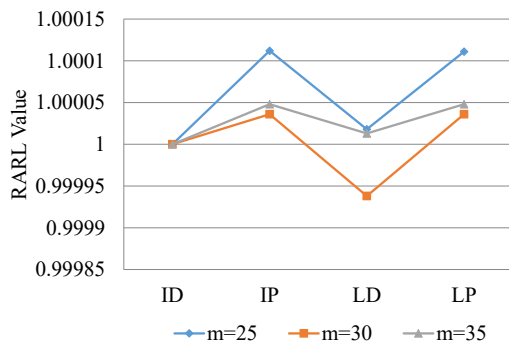


Fig. 14. Relative average run length of exponentially weighted moving average values at *n*=5 of identity and log link functions for both deviance and Pearson residuals

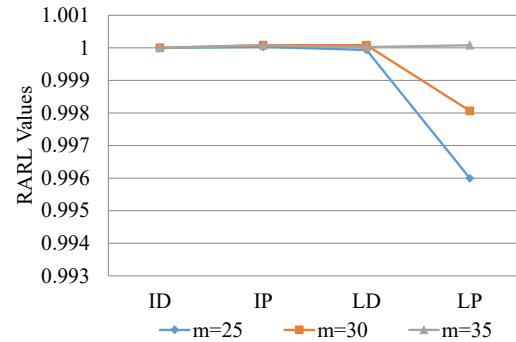


Fig. 15. Relative average run length of exponentially weighted moving average values at *n*=10 of identity and log link functions for both deviance and Pearson residuals

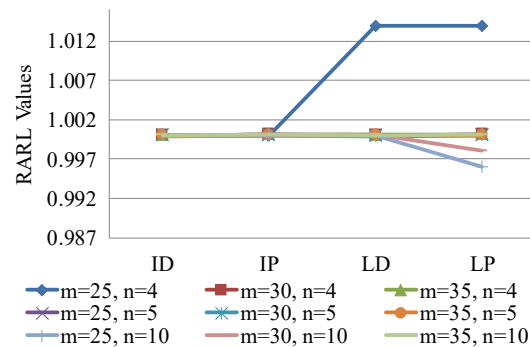


Fig. 16. Overall relative average run length values for all link functions under different types of residuals

From Fig. 15 it can be concluded that, at *m*=25 there were the best values of EWMA for Pearson residuals under log link functions, then at *m*=30 became the second-best values of EWMA for Pearson residuals under log link function.

From Fig. 16 it can be concluded that, at *m*=25 there were the best values of EWMA for Pearson residuals under log link functions, then at *m*=30 became the worst values of EWMA for Pearson residuals under log link function.

## 6. Discussion of the results of profile monitoring of residuals control charts under gamma regression models

The control chart monitors the average, or the centering of the distribution of data from the process. The bottom chart monitors the range, or the width of the distribution. Control charts for attribute data are used singly. To facilitate the comparison between different methods of estimating, let's use RARL, Table 2 showed the relative performance of ARL for all ID, IP, LD, and LP values relative to ID value. It's also the performance of the residual control charts was also monitored by figures.

From Fig. 14, it can be concluded that, at *m*=30 there were the best values of EWMA for log link functions and for deviance residuals.

From Fig. 15 it can be concluded that, at *m*=25 there were the best values of EWMA for Pearson residuals under log link functions, then at *m*=30 became the second-best values of EWMA for Pearson residuals under log link function.

From Fig. 16 it can be concluded that, at *m*=25 there were the best values of EWMA for Pearson residuals under log link functions, then at *m*=30 became the worst values of EWMA for Pearson residuals under log link function.

The EWMA chart is a suitable alternative to other charts for detecting inaccuracy, especially where minor shifts are of concern. It gives a flexible tool for depicting imprecision and inaccuracy. However, detecting imprecision using EWMA charts need for specialized adjustment. [22] compared the effectiveness of the Shewhart  $\bar{x}$  EWMA, residual control charts for auto-correlation observations, and the comparison of the control charts was based on the ARL also.

One of the limitations inherent in the specific research is its application to large sample sizes. Most previous research confirmed that profile monitoring control charts give the best results for a sample of size four or five. An intense comparison could have been carried out by examining the performance of other types of residuals such as standardized residuals and working residuals and by using various measures of performance for evaluation as average time to signal and average number of units sampled until signal under other fitting methods such as Bayesian approach and weighted least square method.

Some results were unsatisfactory with identity and log link functions used in the case of a large sample size, so it can be suggested to find a new link function, but it may be difficult mathematically. Also, the bootstrap method that is scientifically proven to improve the results of the performance estimators may be proposed to use.

Generally, and from Table 2 it can be concluded that the deviance residuals results are more accurate than the Pearson residuals results. Also, the different between the two gamma regression models emphasize that EWMA is an effective alternative to the traditional Shewhart control chart when small shifts in the process parameters are monitored which is compatible with [6] and [7] results.

It is worth mentioning regarding the use of GLMs – based control charts for monitoring gamma distributed response variables and considered as one of the advantages of this study is the use of two types of residuals unlike [15] that proposed to monitor only the deviance residuals comparing with the original observations using the classical Shewhart chart. Another advantage of this study is the residuals monitoring using EWMA chart which is more efficient (more rapidly) to detect small and moderate shifts in a process with an equal sample size than Shewhart chart [11, 12, 18, 19].

In our view, the main drawback to this study as previously mentioned is some unsatisfactory results when using a sample size greater than 5. So, it is maybe suitable to recommend proposing a new joint EWMA chart or monitoring other types of residuals to try to overcome this disadvantage.

---

### 7. Conclusions

---

1. For the first gamma model, from the simulation result with the identity function, it can be generally concluded

that the EWMA results for deviance residuals gave same percentages for OC (lower or upper) samples such as EWMA results for Pearson in most of cases. But at the other cases, specifically, at ( $n=10$  &  $m=30$ ) OC samples increased by one sample (lower) for deviance residuals. This is probably due to the greater efficiency of the deviance residuals over the Pearson residuals with the identity link function.

2. For the second gamma model, from this study results, with the log function, EWMA results for deviance residuals were identical with EWMA results for Pearson in approximately 50 % of cases. At the other 50 % of cases number of OC samples increases sometimes with deviance residuals and sometimes with Pearson residuals.

3. At sample size five, the identity link function gave OC samples more than the log link function. And at sample size ten, the identity link function gave OC samples more than the log link function. This may give an indication that using log link function instead of identity link function gives more accurate results with small and large sample sizes.

4. At any number of samples and by increasing the sample size, the percentages of OC samples increased with identity link function. This didn't satisfy with log link function. This also may give an indication that using log link function instead of identity link function gives more accurate results with any number of samples. Accordingly, it can be said that both flexibility and usefulness are achieved with using the log link function. Usually, the choice of link function is depending on the problem and the data at hand.

---

### Conflict of interest

---

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

---

### Financing

---

The study was performed without financial support.

---

### Data availability

---

Manuscript has no associated data.

---

### Acknowledgments

---

We express our gratitude to Faculty of Graduate Studies for Statistical Research – Cairo University.

---

### References

1. Mahmoud, M. A. (2011). Simple Linear Profiles. *Statistical Analysis of Profile Monitoring*, 21–92. doi: <https://doi.org/10.1002/9781118071984.ch2>
2. Mahmoud, M. A. (2012). The Performance of Phase II Simple Linear Profile Approaches when Parameters Are Estimated. *Communications in Statistics - Simulation and Computation*, 41 (10), 1816–1833. doi: <https://doi.org/10.1080/03610918.2011.621570>
3. Amiri, A., Koosha, M., Azhdari, A., Wang, G. (2014). Phase I monitoring of generalized linear model-based regression profiles. *Journal of Statistical Computation and Simulation*, 85 (14), 2839–2859. doi: <https://doi.org/10.1080/00949655.2014.942864>
4. Noorossana, R., Eyvazian, M., Vaghefi, A. (2010). Phase II monitoring of multivariate simple linear profiles. *Computers & Industrial Engineering*, 58 (4), 563–570. doi: <https://doi.org/10.1016/j.cie.2009.12.003>



5. Mahmoud, M. A., Saad, A. E. N., El Shaer, R. (2014). Phase II Multiple Linear Regression Profile with Small Sample Sizes. *Quality and Reliability Engineering International*, 31 (5), 851–861. doi: <https://doi.org/10.1002/qre.1644>
6. Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1 (3), 239–250. doi: <https://doi.org/10.1080/00401706.1959.10489860>
7. Crowder, S. V. (1987). A Program for the Computation of ARL for Combined Individual Measurement and Moving Range Charts. *Journal of Quality Technology*, 19 (2), 103–106. doi: <https://doi.org/10.1080/00224065.1987.11979046>
8. Lucas, J. M., Saccucci, M. S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32 (1), 1–12. doi: <https://doi.org/10.1080/00401706.1990.10484583>
9. Woodall, W. H., Spitzner, D. J., Montgomery, D. C., Gupta, S. (2004). Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology*, 36 (3), 309–320. doi: <https://doi.org/10.1080/00224065.2004.11980276>
10. Knoth, S. (2007). Accurate ARL Calculation for EWMA Control Charts Monitoring Normal Mean and Variance Simultaneously. *Sequential Analysis*, 26 (3), 251–263. doi: <https://doi.org/10.1080/07474940701404823>
11. Maravelakis, P. E., Castagliola, P., Khoo, M. B. C. (2017). Run length properties of run rules EWMA chart using integral equations. *Quality Technology & Quantitative Management*, 16 (2), 129–139. doi: <https://doi.org/10.1080/16843703.2017.1372853>
12. Khan, N., Yasmin, T., Aslam, M., Jun, C.-H. (2018). On the performance of modified EWMA charts using resampling schemes. *Operations Research and Decisions*, 3, 29–43. doi: <https://doi.org/10.5277/ord180303>
13. Akram, M. N., Amin, M., Qasim, M. (2021). A new biased estimator for the gamma regression model: Some applications in medical sciences. *Communications in Statistics – Theory and Methods*, 50 (23), 1–21. doi: <https://doi.org/10.1080/03610926.2021.1977958>
14. Krishnamoorthy, K. (2006). *Handbook of statistical distributions with applications*. Chapman and Hall/CRC, 376. doi: <https://doi.org/10.1201/9781420011371>
15. Jearkpaporn, D., Montgomery, D. C., Runger, G. C., Borror, C. M. (2003). Process monitoring for correlated gamma-distributed data using generalized-linear-model-based control charts. *Quality and Reliability Engineering International*, 19 (6), 477–491. doi: <https://doi.org/10.1002/qre.521>
16. Braimah, O. J., Osanaiye, P. A., Omaku, P. E., Saheed, Y. K., Eshimokhai, S. A. (2014). On the Use of Exponentially Weighted Moving Average (Ewma) Control Chart in Monitoring Road Traffic Crashes. *International Journal of Mathematics and Statistics Invention (IJMSI)*, 2 (5), 01–09. Available at: <https://ijmsi.org/Papers/Volume.2.Issue.5/A0250109.pdf>
17. Cepeda-Cuervo, E., Corrales, M., Cifuentes, V., Zarate, H. (2016). On Gamma Regression Residuals. *Journal of The Iranian Statistical Society*, 15 (1), 29–44. Available at: [http://jirss.irstat.ir/browse.php?a\\_code=A-11-304-2&slc\\_lang=en&sid=1](http://jirss.irstat.ir/browse.php?a_code=A-11-304-2&slc_lang=en&sid=1)
18. Lu, C.-W., Reynolds, M. R. (1999). EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology*, 31 (2), 166–188. doi: <https://doi.org/10.1080/00224065.1999.11979913>
19. Wang, H. (2016). Application of Residual-Based EWMA Control Charts for Detecting Faults in Variable-Air-Volume Air Handling Unit System. *Journal of Control Science and Engineering*, 2016, 1–7. doi: <https://doi.org/10.1155/2016/1467823>
20. Kang, L., Albin, S. L. (2000). On-Line Monitoring When the Process Yields a Linear Profile. *Journal of Quality Technology*, 32 (4), 418–426. doi: <https://doi.org/10.1080/00224065.2000.11980027>
21. Messaoud, A., Theis, W., Hering, F., Weihs, C. (2008). Monitoring a Drilling Process Using Residual Control Charts. *Quality Engineering*, 21 (1), 1–9. doi: <https://doi.org/10.1080/08982110802355869>
22. Areepong, Y. (2013). A comparison of performance of residual control charts for trend stationary AR(p) processes. *International Journal of Pure and Applied Mathematics*, 85 (3). doi: <https://doi.org/10.12732/ijpam.v85i3.13>
23. García-Bustos, S., Cárdenas-Escobar, N., Debón, A., Pincay, C. (2021). A control chart based on Pearson residuals for a negative binomial regression: application to infant mortality data. *International Journal of Quality & Reliability Management*, 39 (10), 2378–2399. doi: <https://doi.org/10.1108/ijqrm-03-2021-0062>
24. Jamal, A., Mahmood, T., Riaz, M., Al-Ahmadi, H. M. (2021). GLM-Based Flexible Monitoring Methods: An Application to Real-Time Highway Safety Surveillance. *Symmetry*, 13 (2), 362. doi: <https://doi.org/10.3390/sym13020362>
25. Kinat, S., Amin, M., Mahmood, T. (2019). GLM-based control charts for the inverse Gaussian distributed response variable. *Quality and Reliability Engineering International*, 36 (2), 765–783. doi: <https://doi.org/10.1002/qre.2603>
26. Nelder, J. A., Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)*, 135 (3), 370. doi: <https://doi.org/10.2307/2344614>
27. McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall/CRC. doi: <https://doi.org/10.1007/978-1-4899-3242-6>
28. Bossio, M. C., Cuervo, E. C. (2015). Gamma regression models with the Gammareg R package. *Comunicaciones En Estadística*, 8 (2), 211. doi: <https://doi.org/10.15332/s2027-3355.2015.0002.05>
29. Cuervo, E. C. (2001). *Modelagem da Variabilidade em Modelos Lineares Generalizados*. Instituto de Matemáticas, Universidade Federal do Rio de Janeiro.
30. Gipe, G. W. (1976). Using residual analysis to search for specification errors. *Decision Sciences*, 7 (1), 40–56. doi: <https://doi.org/10.1111/j.1540-5915.1976.tb00656.x>
31. Shao, Y. E., Lin, Y. (2013). Applying Residual Control Charts to Identify the False Alarms in a TFT-LCD Manufacturing Process. *Applied Mathematics & Information Sciences*, 7 (4), 1459–1464. doi: <https://doi.org/10.12785/amis/070426>
32. Pregibon, D. (1981). Logistic Regression Diagnostics. *The Annals of Statistics*, 9 (4). <https://doi.org/10.1214/aos/1176345513>
33. Cordeiro, G. M., Simas, A. B. (2009). The distribution of Pearson residuals in generalized linear models. *Computational Statistics & Data Analysis*, 53 (9), 3397–3411. doi: <https://doi.org/10.1016/j.csda.2009.02.025>
34. Pierce, D. A., Schafer, D. W. (1986). Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, 81 (396), 977–986. doi: <https://doi.org/10.1080/01621459.1986.10478361>