JIRSS (2016)
Vol. 15, No. 1, pp 29-44
DOI: 10.7508/jirss.2016.01.002

On Gamma Regression Residuals

Edilberto Cepeda-Cuervo, Martha Corrales, Maria Victoria Cifuentes, Hector Zarate

Departamento de Estadística, Universidad Nacional de Colombia

Abstract. In this paper, we propose new residuals for gamma regression models, assuming that both mean and shape parameters follow regression structures. The models are summarized and fitted by applying both classic and Bayesian methods as proposed by Cepeda-Cuervo (2001). The residuals are proposed from properties of the biparametric exponential family of distributions. Simulated and real data sets are analyzed to determine the performance and behavior of the proposed residuals.

Keywords. Bayesian estimation, Gamma regression, Fisher scoring algorithm, Residuals.

MSC: 62J12; 62J05.

1 Introduction

The gamma distribution can be used for regression models with more flexibility than other models, such as the exponential and Poisson, among others. Thus, gamma regression models allow for a monotone, no constant hazard in survival models, and have the reproductive property that the sums of independent gamma distributions are also gamma distributed. Moreover, gamma regression models have the advantage of providing a count-data model with substantially higher flexibility than the Poisson model, which involves very sparse time-series that can be modeled by the gamma regression (Bateson , 2009). These models are used in a wide range of empirical applications, such as the process of rate setting in the frame-work of heterogeneous insurance portfolios (Krishnamoorthy , 2006) and in hospital admissions for rare diseases where time series are very sparse due to infrequency of events (Winklemann , 2008).

Edilberto Cepeda-Cuervo(✉)(Corresponding Author: ecepedac@unal.edu.co), Martha Corrales (mlcorralesb@unal.edu.co), Maria-Victoria Cifuentes (mvcifuentesa@unal.edu.co), Hector Zarate (hmzarates@unal.edu.co).

This paper considers gamma regression models in which both the mean and the dispersion are allowed to depend on unknown parameters and on covariates. Joint modeling of the mean and the shape parameters in gamma regressions were proposed by Cepeda-Cuervo (2001), under a classic method and a Bayesian approach. In the former, he proposed an alternative iterated maximum likelihood method based on the Fisher scoring algorithm for the parameter estimation and, in the Bayesian approach, he proposed an hybrid Metropolis Hasting algorithm for the regression parameter estimation.

Several definitions of residuals are possible for generalized linear models (McCullagh and Nelder , 1989). Some uses of generalized residuals include: building goodness of fit measures to check for systematic departure from the model, checking the variance function and the link function, examining them to identify poorly fitting observations, and plotting them to examine effects of new covariates or nonlinear effects of the covariates included in the model. Some of the relevant contributions related to residuals in generalized linear models are presented in Cox and Snell (1968), Pierce and Schafer (1986) and Dobson (2010).

In this paper we propose and adjust two residuals for gamma regression models. Simulated and real data applications are used to evaluate the benefits and interpretation of the proposed residuals.

This paper consists of six sections. In Section 2, a re-parameterization of the gamma distribution is presented. In Section 3, the gamma regression models are defined, and both classic and Bayesian methods used to fit these models are summarized. Section 4 presents the residuals obtained under the two-parameter exponential family of distributions. Section 5 contains two applications based on simulated and real data. In that section, we mention two application cases: the first one is based on simulated gamma data which is useful to evaluate residuals' behavior, whereas the second application uses data from a study presented in McCullagh and Nelder (1989) related to the duration of embryonic stage in fruit flies, and where we calculate the gamma residuals to measure adjustment of the model proposed by the authors. Finally, in Section 6, we present our main conclusions.

2 Re-parameterized gamma distribution

A random variable Y follows a gamma distribution if its density function is given by

$$f(y;\lambda,\alpha)=\frac{\lambda}{\Gamma(\alpha)}(\lambda y)^{\alpha-1}e^{-\lambda y}I_{(0,\infty)}(y), \tag{2.1}$$

where $\lambda > 0$, $\alpha > 0$, $\Gamma(.)$ is the gamma function and $I(.)$ is an indicator function. Under this parameterization, the mean and variance of Y are given by $\mu = E(Y) = \alpha/\lambda$ and $\text{Var}(Y) = \alpha/\lambda^2 = \mu^2/\alpha$, respectively.

Setting $\lambda = \alpha/\mu$, Cepeda-Cuervo (2001) and Cepeda and Gamerman (2005) write the gamma density function (2.1) in terms of the mean and shape parameters as

$$f(y)=\frac{1}{y\Gamma(\alpha)}\left(\frac{\alpha y}{\mu}\right)^{\alpha}e^{-\alpha y/\mu}I_{(0,\infty)}(y). \tag{2.2}$$

Under this re-parameterization, we use $Y \sim G(\mu, \alpha)$ to denote that the random variable Y follows a gamma distribution with $E(Y) = \mu$ and α as the shape parameter. The variance of Y is now given by $Var(Y) = \mu^2/\alpha$.

3 Gamma regression models

Let $Y_i \sim G(\mu_i, \alpha), i = 1, \ldots, n,$ be independent random variables. Then the gamma regression model is defined as

$$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown regression parameters ($p < n$), $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})'$ is the vector of p covariates, and η_i is a linear prediction. Usually $x_{i1} = 1$ for all i. So the model has a mean intercept. The link function $g(.) : (0, \infty) \rightarrow \mathbb{R}$ should be a strictly monotonic twice differentiable function in the classic approach and once differentiable in the Bayesian approach.

Some usual link functions in the gamma regression are log ($g(\mu) = log(\mu)$), identity ($g(\mu) = \mu$), and inverse ($g(\mu) = 1/\mu$). In the exponential family, the canonical link for the mean is the inverse function (McCullagh and Nelder , 1989).

An extension of the gamma regression model presented in McCullagh and Nelder (1989) was proposed in Cepeda-Cuervo (2001), assuming that both mean and shape parameters follow regression structures. He further assumed that $Y_i \sim G(\mu_i, \alpha_i)$, $i = 1, \ldots, n,$ are independent random variables with gamma distribution, where the mean and shape parameters follow a regression structure given by

$$g(\mu_i) = \eta_{1i} = \boldsymbol{x_i'\beta}, \tag{3.2}$$

$$h(\alpha_i) = \eta_{2i} = \boldsymbol{z_i'\gamma}, \tag{3.3}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$, $p + k < n$, are the vectors of regression parameters related to the mean and dispersion, respectively, g is the mean link function, h is the shape link function (usually the log function), η_{1i} and η_{2i} are the linear predictors, and $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ are the *ith* observed values of the covariates.

3.1 Classic estimation

Cepeda-Cuervo (2001) proposed a classic approach to fit joint mean and shape gamma regression models using the Fisher scoring algorithm. In that work, he showed that, under the reparameterization of the gamma distribution given by (2.2), the likelihood function of the gamma regression models defined by (3.2) and (3.3) is given by

$$L = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha_i)} \left(\frac{\alpha_i}{\mu_i}\right)^{\alpha_i} y_i^{\alpha_i - 1} \exp\left(-\frac{\alpha_i}{\mu_i} y_i\right) \tag{3.4}$$

and the log likelihood function is given by

$$l = \sum_{i=1}^{n} \left\{ -\log[\Gamma(\alpha_i)] + \alpha_i \log\left(\frac{\alpha_i y_i}{\mu_i}\right) - \log(y_i) - \left(\frac{\alpha_i}{\mu_i}\right) y_i \right\}. \tag{3.5}$$

Thus, assuming that $\mu_i = \mathbf{x_i'}\boldsymbol{\beta}$ and $\alpha_i = \exp(\mathbf{z_i'}\boldsymbol{\gamma})$, the components of the score function are as

$$\begin{aligned}
\frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} -\frac{\alpha_i}{\mu_i}\left(1 - \frac{y_i}{\mu_i}\right) x_{ij}; \quad j = 1, \ldots p, \\
\frac{\partial l}{\partial \gamma_k} &= \sum_{i=1}^{n} -\alpha_i \left[\frac{d}{d\alpha_i} \log\Gamma(\alpha_i) - \log\left(\frac{\alpha_i y_i}{\mu_i}\right) - 1 + \frac{y_i}{\mu_i}\right] z_{ik}; \quad k = 1, \ldots, r.
\end{aligned}$$

On the other hand, the Hessian matrix is determined by

$$\begin{aligned}
\frac{\partial^2 l}{\partial \beta_k \beta_j} &= \sum_{i=1}^{n} \frac{\alpha_i}{\mu_i^2}\left(1 - \frac{2y_i}{\mu_i}\right) x_{ij} x_{ik}; \quad j,k = 1, \ldots p, \\
\frac{\partial^2 l}{\partial \gamma_k \beta_j} &= \sum_{i=1}^{n} -\frac{\alpha_i}{\mu_i}\left(1 - \frac{y_i}{\mu_i}\right) x_{ij} z_{ik}; \quad k = 1, \ldots, r, \\
\frac{\partial^2 l}{\partial \gamma_k \gamma_j} &= \sum_{i=1}^{n} -\alpha_i \left[\frac{d}{d\alpha_i} \log\Gamma(\alpha_i) - \log\left(\frac{\alpha_i y_i}{\mu_i}\right) - 1 + \frac{y_i}{\mu_i}\right] z_{ij} z_{ik}, \\
&\quad - \sum_{i=1}^{n} \alpha_i \left[\alpha_i \frac{d^2}{d\alpha_i^2} \log\Gamma(\alpha_i) - 1\right] z_{ij} z_{ik}; \quad j,k = 1, \ldots, r.
\end{aligned}$$

The Fisher information matrix is given by

$$\begin{aligned}
-E\left(\frac{\partial^2 l}{\partial \beta_k \beta_j}\right) &= \sum_{i=1}^{n} \frac{\alpha_i}{\mu_i^2} x_{ji} x_{ki}; \quad j,k = 1, ..., p, \\
-E\left(\frac{\partial^2 l}{\partial \gamma_k \beta_j}\right) &= 0; \quad k = 1, ..., r; \quad j = 1, ..., p, \\
-E\left(\frac{\partial^2 l}{\partial \beta_k \beta_j}\right) &= \sum_{i=1}^{n} \alpha_i^2 \left[\frac{d^2}{d\alpha_i^2} \log\Gamma(\alpha_i) - \frac{1}{\alpha_i}\right] z_{ij} z_{ki}; \quad j,k = 1, ..., r
\end{aligned}$$

It can be noted that the Fisher information matrix is a block diagonal matrix, where one of the blocks corresponds to the mean regression parameters and the other to the shape regression parameters. Thus the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are orthogonal, and their maximum likelihood estimators, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}$, are asymptotically independent. As a consequence of this result, Cepeda-Cuervo (2001) proposed an iterative algorithm to obtain the maximum likelihood estimates of the regression parameters, where, given the k-th parameter values $(\boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)})'$, the mean vector of the regression parameters is updated from

$$\boldsymbol{\beta}^{(k+1)} = (\boldsymbol{X}'\boldsymbol{W}^{(k)}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(k)}\boldsymbol{Y}, \tag{3.6}$$

where $\boldsymbol{W}^{(k)}$ is a matrix with diagonal elements $w_i^{(k)} = (\mu_i^2)^{(k)}/\alpha_i^{(k)}$, and, given $(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\gamma}^{(k)})'$, the shape regression parameters $\boldsymbol{\gamma}^{(k+1)}$ are updated from the equation

$$\boldsymbol{\gamma}^{(k+1)} = (\boldsymbol{Z}'\boldsymbol{W}^{(k)}\boldsymbol{Z})^{-1}\boldsymbol{X}'\boldsymbol{W}^{(k)}\boldsymbol{Y}, \tag{3.7}$$

where $\boldsymbol{W}^{(k)}$ is a matrix with elements $w_i^{(k)} = 1/d_i^{(k)}$ in which

$$d_i = \alpha_i^{-2}\left[\frac{d^2}{d\alpha_i^2}\log\Gamma(\alpha_i) - \frac{1}{\alpha_i}\right]^{-1}. \tag{3.8}$$

Therefore, given the initial values of the parameters, an alterative iterate algorithm can be summarized as follows.

1. Give initial values for the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.
2. Obtain $\boldsymbol{\beta}^{(k+1)}$ from equation (3.6).
3. Obtain $\boldsymbol{\gamma}^{(k+1)}$ from equation (3.7).
4. Return to 2 until convergence.

3.2 Bayesian estimation

In this section, we summarize the Bayesian method proposed by Cepeda-Cuervo (2001) to fit gamma regression models, where both mean and shape parameters follow regression structures. In this proposal, without loss of generality, independent normal prior distributions are assumed for mean and shape regression parameters, that is,

$$\begin{aligned}\boldsymbol{\beta} &\sim N(\mathbf{b}, \mathbf{B}),\\ \boldsymbol{\gamma} &\sim N(\mathbf{g}, \mathbf{G}).\end{aligned}$$

Let $L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{Y}, \mathbf{X}, \mathbf{Z})$ be the likelihood function and $p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ be the joint prior distribution. Given that the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}|\mathbf{Y}, \mathbf{X}, \mathbf{Z}) \times$
$p(\boldsymbol{\beta}, \boldsymbol{\gamma})$ and all their conditional distributions $\pi_\beta(\boldsymbol{\beta}|\boldsymbol{\gamma}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ and $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ are analytically intractable, an alternate Metropolis Hastings algorithm is proposed to obtain samples of the posterior parameters.

In this algorithm, samples of the conditional posterior distribution $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ are proposed from the kernel transition function, which is given by

$$q_1(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = N(\mathbf{b}^*, \mathbf{B}^*), \tag{3.9}$$

where

$$\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\widetilde{Y}),$$
$$\mathbf{B}^* = (\mathbf{B}^{-1} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}.$$

For identity and log mean link functions, the components of the working variables $\widetilde{Y}$ are $\widetilde{y_1}_i = y_i$ and $\widetilde{y_1}_i = \mathbf{x}_i'\beta + y_i/\mu_i - 1$, respectively. $\boldsymbol{\Sigma}$ is a diagonal matrix with $w_i = Var(\widetilde{y_1}_i)$, $i = 1, \ldots, n$, as diagonal elements.

Samples of the posterior conditional distribution $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \mathbf{Y}, \mathbf{X}, \mathbf{Z})$ are proposed from the kernel transition function

$$q_2(\gamma|\hat{\beta}, \hat{\gamma}) = N(\mathbf{g}*, \mathbf{G}*), \tag{3.10}$$

where

$$\mathbf{g}^* = \mathbf{G}^*(\mathbf{G}^{-1}\mathbf{g} + \mathbf{X}'\Psi^{-1}\widetilde{Y}),$$
$$\mathbf{G}^* = (\mathbf{G}^{-1} + \mathbf{X}'\Psi^{-1}\mathbf{X})^{-1}.$$

For log link function for the shape, the working variable is $\tilde{y}_{2i} = \mathbf{z}_i'\gamma + y_i/\mu_i - 1$. $\boldsymbol{\Psi}$ is a diagonal matrix with $\varphi_i = Var(\widetilde{y_{2i}})$, $i = 1, \ldots, n$.

For more details about this algorithm and its applications, see Cepeda-Cuervo (2001) and Cepeda and Gamerman (2005).

Having the kernel transition functions defined by (3.9) and (3.10), the hybrid Metropolis Hasting algorithm is defined by the following steps.

1. Begin the chain iteration counter at j=1.
2. Set initial chain values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively.
3. Propose a new value $\boldsymbol{\phi}$ for $\boldsymbol{\beta}$, generated from 3.9.
4. Calculate the acceptance probability, $\alpha(\boldsymbol{\beta}^{(j-1)}, \boldsymbol{\phi})$. If the movement is accepted, then $\boldsymbol{\beta}^{(j)} = \boldsymbol{\phi}$. If not accepted, then $\boldsymbol{\beta}^{(j)} = \boldsymbol{\beta}^{(j-1)}$.
5. Propose a new value $\boldsymbol{\phi}$ for $\boldsymbol{\gamma}$, generated from 3.10.
6. Calculate the acceptance probability $\alpha(\boldsymbol{\gamma}^{(j-1)}, \boldsymbol{\phi})$. If the movement is accepted, then $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\phi}$. If it is not accepted, then $\boldsymbol{\gamma}^{(j)} = \boldsymbol{\gamma}^{(j-1)}$.
7. Change the counter from j to $j + 1$ and return to 2 until convergence is reached.

The convergence can be verified empirically in different ways (for details see Gamerman and Lopes (2006) and Heidelberger and Welch (1981)).

4 Gamma regression residuals

Residual analysis aims to identify outliers and/or model misspecification. It can be based on ordinary residuals, standardized variants or deviance residuals. Residuals are measures of agreement between the observed responses and the fitted conditional mean. Most residuals are based on the differences between the observed responses and the fitted conditional mean. For the gamma regression, where both mean and shape parameters follow regression structures, we define a first standardized ordinal residual as

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{Var}(y_i)}}, \tag{4.1}$$

where

$$\widehat{Var}(y_i) = \frac{\hat{\mu}_i^2}{\hat{\alpha}_i}. \tag{4.2}$$

A second residual considered in this paper is the deviance residual, which for gamma regression models is given by

$$r_i^d = -2\sum_{i=1}^{n}\left[\log\left(\frac{y_i}{\widehat{\mu_i}}\right) - \frac{y_i - \widehat{\mu_i}}{\widehat{\mu_i}}\right], \tag{4.3}$$

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

In order to define gamma residuals from the two parameter exponential family we re-parameterized the gamma density function in a natural way, as follows in equation (4.6), where $\eta_1 = \alpha, T_1 = \log(y), \eta_2 = -\frac{\alpha}{\mu}, T_2 = y, d_0(\eta_1, \eta_2) = \eta_1 \log(\eta_2) - \log\Gamma(\eta_1), S(y) = -\log(y)$.

$$\begin{aligned} f(y) &= \exp\left[-\log\Gamma(\alpha) + \alpha\log\left(\frac{\alpha y}{\mu}\right) - \frac{\alpha y}{\mu} - \log(y)\right] && (4.4)\\ &= \exp\left[\alpha\log(y) - \Big(\frac{\alpha}{\mu}\Big)y + \alpha\log\Big(\frac{\alpha}{\mu}\Big) - \log\Gamma(\alpha) - log(y)\right] && (4.5)\\ &= \exp\left[\eta_1 T_1(y) + \eta_2 T_2(y) + \eta_1\log(-\eta_2) - \log\Gamma(\eta_1) + S(y).\right]. && (4.6)\end{aligned}$$

Thus, from the properties of the bi-parametric exponential family of distributions,

$$\begin{aligned} E(T_1) &= -\frac{\partial d_0}{\partial \eta_1} = -[\log(-\eta_2) - \Psi(\eta_1)], && (4.7)\\ E(T_2) &= -\frac{\partial d_0}{\partial \eta_2} = -\frac{\eta_1}{\eta_2} = \mu, && (4.8)\end{aligned}$$

where the digamma function, $\Psi(\eta_1)$, is defined as the derivative of the logarithm of the gamma function

$$\Psi(\eta_1) = \frac{d\log\Gamma(\eta_1)}{d\eta_1} = \frac{\Gamma'(\eta_1)}{\Gamma(\eta_1)}. \tag{4.9}$$

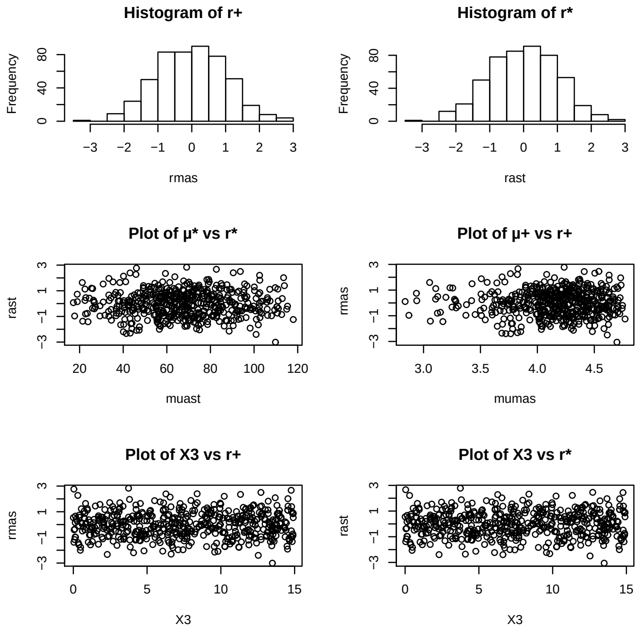

Figure 1: Residuals r+ and r*

From the same properties of the biparametric exponential family, the variances of T_1 and T_2 are given by

$$Var(T_1) = -\frac{\partial^2 d_0}{\partial \eta_1^2} = \Psi'(\eta_1), \quad (4.10)$$

$$Var(T_2) = -\frac{\partial^2 d_0}{\partial \eta_2^2} = \frac{\eta_1}{\eta_2^2} = \frac{\mu^2}{\alpha}, \quad (4.11)$$

where $\Psi'(\eta_1)$ denotes the derivative of the digamma function estimated on η_1

From this result, two gamma residuals can be proposed. The first one from (4.7) and (4.10) is given by

$$r_i^* = \frac{y_i^* - \widehat{\mu_i^*}}{\sqrt{\widehat{Var}(y_i^*)}}, \quad (4.12)$$

where $y_i^* = T_1(y_i) = log(y_i)$, $\mu_i^* = E(T_1(y_i)) = E(y_i^*)$ and $Var(y_i^*) = Var(T_1(y_i)) = \Psi'(\eta_1)$. This residual is computed as the difference between y^* and $\hat{\mu}^*$,the difference between y^* and the

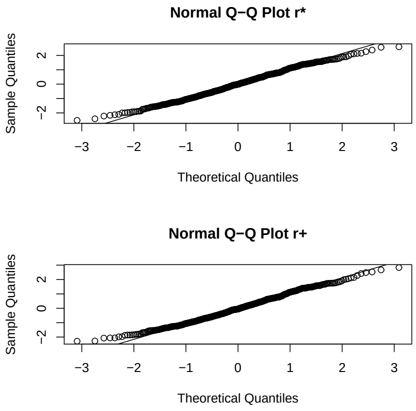

Figure 2: Normal Residuals for r+ and r*

estimates of the expected value $\mu^* = E(y^*)$, divided by the squared root of the estimation of the variance $Var(y^*)$.

Now from (4.8) and (4.11), a second residual can be defined, in this case given by:

$$r_i^+ = \frac{y_i^+ - \hat{\mu}_i^+}{\sqrt{\widehat{Var}(y_i^+)}} \tag{4.13}$$

where $y_i^+ = T_2(y_i) = y_i$, $\mu_i^+ = E(T_2(y_i)) = E(y_i^+)$ and $Var(y_i^+) = Var(T_2(y_i)) = \mu_i^2/\alpha_i)$. This residual is the same as the ordinary standardized residual, but it is obtained from the properties of the two-parameter exponential family of distributions, as in Lehmann and Casella (1998).

5 Applications

In this section, we present two applications. The first one based on the simulated data and the second one using data on the duration of the embryonic stage of fruit flies reported by Powsner (1935) and McCullagh and Nelder (1989).

5.1 Simulation data set

500 values of three covariates were simulated from uniform distributions. Values of the covariates X_2, X_3 and X_4 were generated from uniform distributions $U(0, 30)$, $U(0, 15)$ and $U(10, 20)$, respectively. Values of the covariate X_1 are assumed to be a vector of ones, in order to define mean and shape models with intercept. Values of the response variables Y were generated from a gamma distribution with mean and shape parameters given by

$$\begin{aligned} \hat{\mu}_i &= 15 + 2x_{2i} + 3x_{3i}, & (5.1)\\ \hat{\alpha}_i &= \exp(0.2 + 0.1x_{2i} + 0.3x_{4i}). & (5.2) \end{aligned}$$

The fitted mean equation and the fitted shape equations, obtained by applying the Bayesian method proposed by Cepeda-Cuervo (2001), are given as

$$\begin{aligned} \hat{\mu}_i &= 15.015 + 2.001x_{2i} + 2.998x_{3i}, & (5.3)\\ \hat{\alpha}_i &= \exp(0.360 + 0.104x_{2i} + 0.290x_{4i}). & (5.4) \end{aligned}$$

We consider residual checks for systematic departure from the model using some informal graphs. From Figure 1, in the second panel, both residuals $r+$ and $r*$ are plotted against the varying mean of the model $\widehat{\mu_i}$. Typical systematic deviations are absent due to the fact that there is neither curvatures in the mean nor a systematic change. According to the third panel, where the residuals are plotted against the linear predictor X_3, we conclude that there is no appearance of a systematic trend.

The normal probability plot in Figure 2 (Q-Q plot for $r+$ and $r*$) suggests a good fit of both residuals $r+$ and $r*$ to the normal distribution. As expected, the analysis of the residual under study did not single out any observation as atypical or yield evidence of lack of fit.

Finally, the third plot is the partial residual plot for the gamma regression model, which is used to assess the form of a predictor and is thus calculated for each predictor. If the scale is satisfactory, the plot should be approximately linear. If not, its form suggests a suitable alternative. According to Figure 3, the X_2 variable should have curvature and X_3 should be linear.

In order to determine the performance of the proposed residual, we compared it with the standardized ordinal residual, using simulation studies. First, we repeated the simulation developed before in this section and found that in all of them the normal Q-Q plots had the same shape as in Figure 2, and the Q-Q-plot of r^+ and r^* are very similar. Next, we generated outliers that were associated with positive residuals to determine which of the standardized residuals are the best to reveal their existence. In this case, the normal Q-Q plots had the same shape as the one shown in Figure 5, which clearly suggests that the standardized ordinal residuals are indeed the best ones. Finally, when we generated outliers that were associated with negative residuals, we found that the proposed residual r^* is the best to determine the existence of these outliers, as can be seen in Figure 4.

5.2 Duration of the embryonic stage of fruit flies

This application is based on an example presented by McCullagh and Nelder (1989). They used a data set collected by Powsner (1935) to measure the effect of temperature on the duration of the development stages of the fruit fly (Drosophila melanogaster). In his experiment, there are

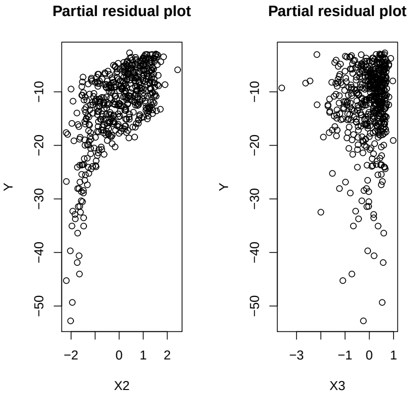

Figure 3: Partial Residuals

four stages: the embryonic, egg-larval, larval and pupal. Only the first is considered here. In this model, observed duration is the response variable, weighted due to batch size.

According to McCullagh and Nelder (1989), the systematic part of the model is considered by rational functions of temperature as

$$\beta_0 + \beta_1 T + \beta_2/(T - \delta), \tag{5.5}$$

where δ represents an asymptote for the temperature function. The fit of the model takes into account the gamma regression and that the identity link was preferred over the inverse and log links, respectively. They adjusted this model considering that the coefficient of variation is constant.

The residuals summarized in this article were calculated by assuming the model

$$\begin{aligned} \mu_i &= \beta_0 + \beta_1 T_i + \beta_2/T_i, & (5.6)\\ \alpha_i &= \exp(\gamma_0 + \gamma_1 T_i + \gamma_2/T_i), & (5.7) \end{aligned}$$

for the fruit fly application, and the following parameter estimates (and standard deviations) were observed: $\hat{\beta}_0 = -2.2828(1,4485)$, $\hat{\beta}_1 = 0.04068(0,0298)$, $\hat{\beta}_2 = 36.7313(17,3253)$, $\hat{\gamma}_0 = 3.3718(2,9484)$, $\hat{\gamma}_1 = -0.0529(0,0671)$ and $\hat{\gamma}_2 = -15.8543(31,0588)$.

In Figure 5, it can be seen that due to the small number of observations, in some panels (such as the fourth one), the residuals (r^*) appear to have linear dependence on μ^*, meaning that, for this case, r^+ is more dependable than r^* in order to get a better residual. Regarding the histograms of r^+ and r^*, we can observe that they are not as accurate as the previous application, which was expected as the first data set was generated by a gamma simulation. However, in this case, the residuals show greater accumulation around zero, but the distribution does not

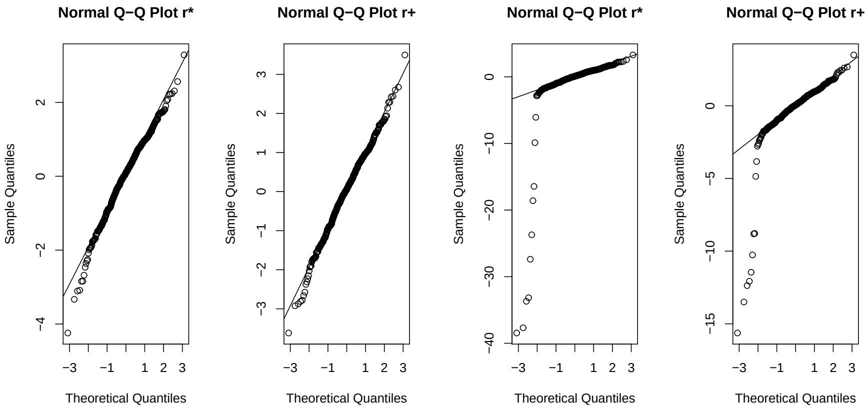

Figure 4: Residual comparison

look symmetric like the normal distribution, possibly because of the relatively small number of observations.

According to the QQ plot in Figure 6, the distribution of both residuals r+ and r* are close to the normal distribution. There is no pattern when we plot the residuals against the covariates.

Finally, Figure 7 summarizes other residuals calculated from the fruit fly data. There are three residuals. The first and second show the estimated μ against the absolute value of each residual (r^+ and r^*). The second new residual was the Pearson residual, which has irregular and scattered behavior, a desirable property in residuals. The last calculated ones are the deviance residuals for both r^+ and r^*, which are shown in panels 5 and 6 in Figure 7.

6 Conclusion

In this paper, we proposed two new residuals for gamma regression models, for which many link functions can be used. We chose the identity and log link for this evaluation. The new residuals were computed by the difference of the link function responses and their fitted means respectively using Fisher scoring and Bayesian estimation of the parameters. The results suggested that the residuals that we call r+ are the same as commonly used ordinary residuals. On the other hand, the new residuals r∗, which come from a Fisher scoring iterative algorithm, were also approximated by the standard normal distribution and fulfill informal checks for systematic departure from the model. This fact can be used to construct more reliable goodness of fit measures and measures of explained variation for gamma regression models.

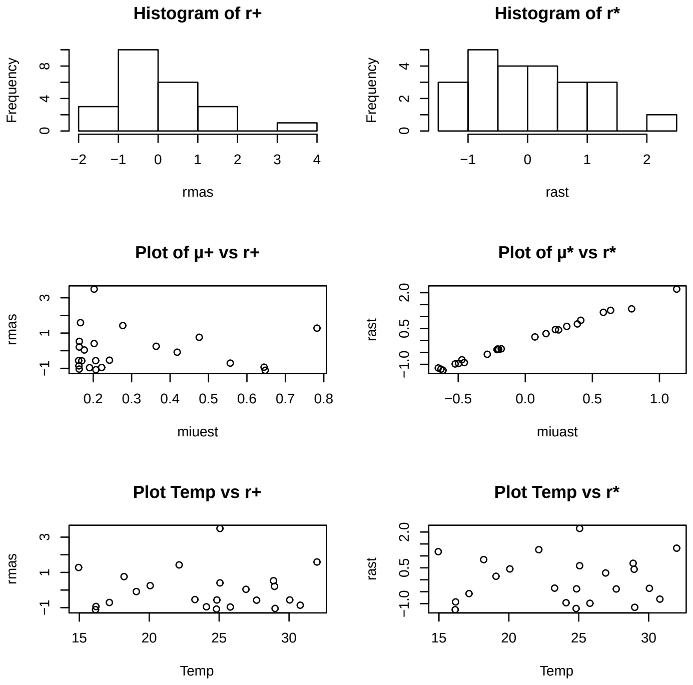

Figure 5: Residuals for r+ and r*

References

Bateson, T. F. (2009), *Gamma Regression of Interevent Waiting Times Versus Poisson Regression of Daily Event Counts: Inside the Epidemiologist's toolboxselecting the Best Modeling Tools for the Job*. Epidemiology, **20**(2), 202-204.

Cepeda-Cuervo, E. (2001), *Modelagem de Variabilidade em Modelos Lineares Generalizados*. Unpublished Ph.D.thesis, Mathematics Institute, Universidade Federal Rio de Janeiro.

Cepeda, E. and Gamerman, D. (2005), *Bayesian Methodology for Modeling Parameters in The Two Parameters Exponential Family*. Estadística, **57**(168), 93-105.

Cox, P.R. and Snell, E.J. (1968), *A General Definition of residuals*. Journal of the Royal Statistical Society, **30**, 248-275.

Gamerman, D. and Lopes, H. F. (2006), *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference*. CRC Press.

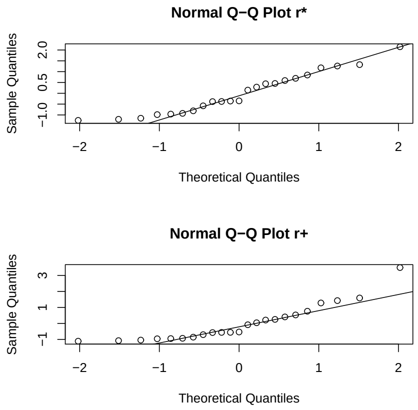

Figure 6: Normality Residuals for r+ and r*

Dobson, A. J. (2010), *An introduction to generalized linear models*. Florida: CRC Press.

Heidelberger, P. and Welch, P. D. (1981), *A spectral method for confidence interval generation and run length control in simulations*. Comm. ACM., **24**(), 233245.

Krishnamoorthy, K. (2006), *Handbook of Statistical Distributions with Applications*. Florida:Chapman & Hall/CRC.

Lehmann, E. L. and Casella, G. (1998), *Theory of point estimation*. New York: Springer.

McCullagh, J. and Nelder, J.A. (1989), *Design and Inference in Finite Population Sampling*. London: Chapman and Hall.

Pierce, D. A. and Schafer, D. W. (1986), *Residuals in Generalized Linear Models*. Journal of the American Statistical Association, **81**(396), 977-986.

Powsner, L. (1935), *The effects of temperature on the durations of the developmental stages of Drosophila melanogaster*. Physiological Zoology, **8**(4), 474-520.

Winklemann, R. (2008), *Econometric analysis of count data*. Berlin, Germany: Springer-Verlag.

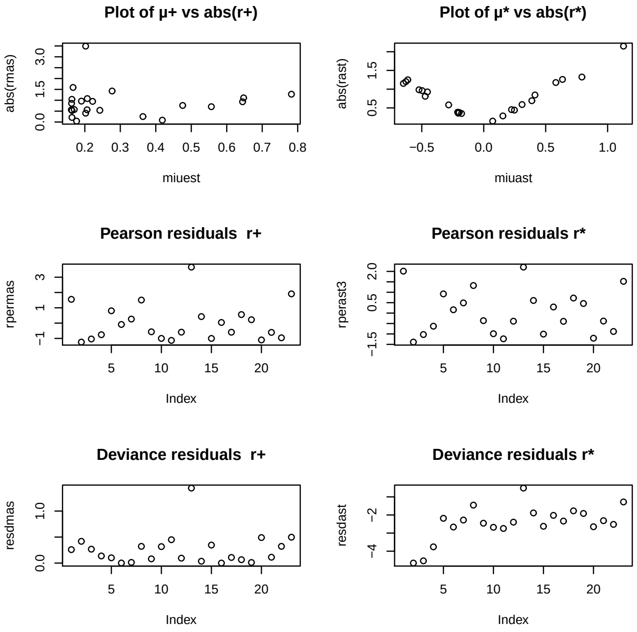

Figure 7: Different Residuals for r+ and r*