

# Generalized Exponentiated Weibull Linear Model in the Presence of Covariates

Tiago V. F. Santana<sup>1</sup>, Edwin M. M. Ortega<sup>2</sup>, Gauss M. Cordeiro<sup>3</sup> & Adriano K. Suzuki<sup>4</sup>

<sup>1</sup> Department of Statistics, Londrina State University, Londrina, Brazil

<sup>2</sup> Department of Mathematical Sciences, University of São Paulo, Piracicaba, Brazil

<sup>3</sup> Department of Statistics, University Federal of Pernambuco, Pernambuco, Brazil

<sup>4</sup> Department of Applied Mathematics and Statistics, University of São Paulo, São Carlos, Brazil

Correspondence: Edwin M. M. Ortega, Av. Pádua Dias 11, 13418-900 – Piracicaba, SP, Brazil, E-mail: edwin@usp.br

Received: August 25, 2016 Accepted: March 28, 2017 Online Published: April 11, 2017

doi:10.5539/ijsp.v6n3p75

URL: <https://doi.org/10.5539/ijsp.v6n3p75>

## Abstract

A new regression model based on the exponentiated Weibull with the structure distribution and the structure of the generalized linear model, called the generalized exponentiated Weibull linear model (GEWLM), is proposed. The GEWLM is composed by three important structural parts: the random component, characterized by the distribution of the response variable; the systematic component, which includes the explanatory variables in the model by means of a linear structure; and a link function, which connects the systematic and random parts of the model. Explicit expressions for the logarithm of the likelihood function, score vector and observed and expected information matrices are presented. The method of maximum likelihood and a Bayesian procedure are adopted for estimating the model parameters. To detect influential observations in the new model, we use diagnostic measures based on the local influence and Bayesian case influence diagnostics. Also, we show that the estimates of the GEWLM are robust to deal with the presence of outliers in the data. Additionally, to check whether the model supports its assumptions, to detect atypical observations and to verify the goodness-of-fit of the regression model, we define residuals based on the quantile function, and perform a Monte Carlo simulation study to construct confidence bands from the generated envelopes. We apply the new model to a dataset from the insurance area.

**Keywords:** Exponentiated Weibull model, generalized linear model, residuals, sensibility.

## 1. Introduction

The Weibull distribution has been widely used in biomedical and industrial studies and to analyze lifetime data in several areas. According to (Collet, 2003), it is just as important for parametric analysis of survival data as the Normal distribution is in linear models. However, the Weibull distribution does not produce good results in situations that present a non-monotone hazard rate function (hrf), such as the unimodal and U-shaped (or bathtub) functions, which are common in studies of reliability, biology and engineering. Due to this problem, (Mudholkar & Srivastava, 1993) proposed the exponentiated Weibull (EW) model, whose hrf can be either monotone or non-monotone. Sub-models of the EW distribution are the Weibull, Exponential, Rayleigh, Burr type X and Exponentiated-Exponential (EE) distributions, among others.

The EW distribution has been widely applied in many fields of knowledge, both practical and theoretical. Researchers in the computer sciences, clinical drug trials, hydrology, silviculture, statistics, actuary, reliability, ecology and mechanical engineering, have employed this distribution successfully. As an example, we can mention the works of (Mudholkar & Hutson, 1996), who modeled flood data on the Floyd river in the American state of Iowa, (Ahmad, Islam & Salam, 2006) on accelerated lifetime tests, (Surlis & D'Ambrosio, 2004) in applications to carbon fiber compounds, (Wang & Rennolls, 2005) to model the diameter of trees in Chinese pine plantations, (Zhang, Xie & Tang, 2005) in a lifetime study of firmware systems, and (Barghout, 2009), who proposed a new order-based reliability prediction model for failure detection, among many others. More recently, (Nadarajah, Cordeiro & Ortega, 2013) presented an extensive review of the EW distribution, with details on its development, several recent results and main applications.

(Prudente & Cordeiro, 2010) proposed a class of regression models following the same concepts of generalized linear models (GLMs), but assuming the Weibull distribution for the response variable. The generalized Weibull linear model (GWLM), as called by the authors, allows a new use of the Weibull distribution in a structure very similar to that of GLMs. The proposed model was defined as a synthesis of three important GLM structures: the random component for the response variable, following a suitably parameterized Weibull distribution; a systematic component specified by a linear function to model a relevant part of the means distribution, called the linear predictor; and a link function to connect

the random and systematic components. Due to the importance of the Weibull and exponentiated Weibull distributions in analyzing data and following the works by (Prudente & Cordeiro, 2010) and (Vanegas, Rondón & Cordeiro, 2013), we propose a new class of regression models based on the EW distribution, called the *generalized exponentiated Weibull linear model* (GEWLM), and we emphasize the GWLM as special case of the GEWLM. The proposed model is very competitive to the GLMs, in particular to the gamma and inverse Gaussian models.

The inferential part was carried out using the asymptotic distribution of the maximum likelihood estimators (MLEs), which, in situations when the sample is small, may present difficult results to be justified. As an alternative to classic analysis we explore the use of the Bayesian method as a feasible alternative. After modeling, it is important to check assumptions in the model and to conduct a robustness study in order to detect influential or extreme observations that can cause distortions in the results of the analysis. Numerous approaches have been proposed in the literature to detect influential or outlying observations.

In this paper, we use two approaches: local influence and Bayesian case influence. Further, we study the residual analysis to assess departures from the error assumptions and to detect outlying observations in the GEWLM.

The paper is organized as follows. In Section 2, we define the GEWLM and examine some of its properties. In Sections 3, we discuss the new GEWLM regression model. In Section 4, we obtain the MLEs and the estimates based on a Bayesian method and provide some results from simulation studies for the GEWLM regression model. The sensitivity analysis based in local influence and Bayesian case influence diagnostics is developed in Section 5. In Section 6, we define the quantile residuals for the new model and study the method to construct a simulation envelope to assess its goodness of fit to a dataset. In Section 7, we fit the new model to a real dataset in the insurance area to demonstrate the performance of the GEWLM in contrast to the GWLM. Finally, in Section 8, we provide some conclusions and suggestions for future research.

**2. The Generalized Exponentiated Weibull Linear Model**

The GWLMs are very effective, as stated by the authors, to model a part of the mean response in terms of a regression equation for positive continuous data and to define the residuals. Many other diagnostic measures can closely follow the theory of GLMs.

Let  $Y_1, \dots, Y_n$  be  $n$  independent random variables, where each  $Y_i$  has the Weibull distribution with shape parameter  $\phi$  and scale parameter  $\alpha_i$ , which varies according to the observations, with cumulative distribution function (cdf) and probability density function (pdf) given by

$$F(y_i; \alpha_i, \phi) = 1 - \exp \left[ - \left( \frac{y_i}{\alpha_i} \right)^\phi \right] \tag{1}$$

and

$$f(y_i; \alpha_i, \phi) = \phi y_i^{-1} \left( \frac{y_i}{\alpha_i} \right)^\phi \exp \left[ - \left( \frac{y_i}{\alpha_i} \right)^\phi \right], \quad \phi, \alpha_i, y_i > 0, \tag{2}$$

respectively. The mean and variance are

$$E(Y_i) = \alpha_i \Gamma_1(\phi) \quad \text{and} \quad \text{Var}(Y_i) = \alpha_i^2 \{ \Gamma_2(\phi) - \Gamma_1^2(\phi) \},$$

where  $\Gamma_j(\phi) = \Gamma(1 + \frac{j}{\phi})$  for  $j \geq 1$ ,  $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$  is the gamma function and  $\Gamma'(p) = \partial \Gamma(p) / \partial p$ .

By re-parameterizing equation (1) according to the relation  $\lambda_i = \alpha_i \exp[\phi^{-1} \Gamma'(2)]$ , proposed by (Cox & Reid, 1987), we obtain the orthogonality between the parameters  $\phi$  and  $\lambda_i$ , i.e.,  $E(\partial^2 \ell(\theta) / \partial \phi \partial \lambda_i) = E(\partial^2 \ell(\theta) / \partial \lambda_i \partial \phi) = 0$ . The advantage of having orthogonal parameters is the asymptotic independence of the MLEs of  $\phi$  and  $\lambda_i$  and the simplification of the algebraic developments, since the expected information matrix is block diagonal. Then, the re-parameterized expressions for the cdf, pdf, expectation and variance of  $Y_i$  are given by

$$F(y_i; \lambda_i, \phi) = 1 - \exp \left[ - \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \right], \tag{3}$$

and

$$f(y_i; \lambda_i, \phi) = \phi y_i^{-1} \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \exp \left[ - \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \right], \quad \phi, \lambda_i, y_i > 0, \tag{4}$$

$$E(Y_i) = \lambda_i c_1(\phi) \quad \text{and} \quad \text{Var}(Y_i) = \lambda_i^2 c_2(\phi),$$

where  $c_1(\phi) = \exp[-\Gamma'(2)/\phi]\Gamma_1(\phi)$  and  $c_2(\phi) = \exp[-2\Gamma'(2)/\phi][\Gamma_2(\phi) - \Gamma_1^2(\phi)]$ .

The GWLM is defined by the pdf (4) for the random component, and by a function  $g$  applied to  $\lambda_i$  such that  $g(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$  is the systematic component, where  $\mathbf{x}_i$  is the vector of covariates associated with the  $i$ th observation and  $\boldsymbol{\beta}$  is an unknown vector of parameters.

By raising equation (3) to the power  $\delta > 0$ , we obtain the EW distribution with scale parameter  $\lambda_i$  and two shape parameters  $\phi$  and  $\delta$ . Henceforth, the re-parameterized EW distribution is denoted by  $Y \sim \text{EW}(\lambda_i, \phi, \delta)$ . The cdf and pdf of  $Y$  are given by

$$F(y_i; \lambda_i, \phi, \delta) = \left\{ 1 - \exp \left[ - \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \right] \right\}^\delta \tag{5}$$

and

$$f(y_i; \lambda_i, \phi, \delta) = \delta \phi y_i^{-1} \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \exp \left[ - \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \right] \left\{ 1 - \exp \left[ - \left( \frac{y_i}{\lambda_i} \right)^\phi e^{\Gamma'(2)} \right] \right\}^{\delta-1}, \tag{6}$$

respectively. The model (6) is referred to as the EW distribution. If  $\delta = 1$  and  $\phi = 1$ , we obtain the Exponential distribution, whereas if  $\phi = 1$ , we have the EE distribution. If  $Y_i$  follows the EW distribution, then the random variable  $Z_i = (Y_i/\lambda_i)^\phi e^{\Gamma'(2)}$  has the EE distribution with unit scale parameter and shape parameter  $\delta$ . The following theorem presents the result in a more general form.

**Theorem:** If  $Y \sim \text{EW}(\lambda, \phi, \delta)$  and  $Z = (Y/\lambda)^\phi e^{\Gamma'(2)}$ , then  $Z \sim \text{EE}(1, \delta)$ .

**Proof:** Let  $t$  be a point in the support of the EW model. Then,

$$\begin{aligned} F_Z(t) &= P(Z \leq t) = P\left(\frac{Y}{\lambda}^\phi e^{\Gamma'(2)} \leq t\right) = P\left(Y^\phi \leq \frac{t \lambda^\phi}{e^{\Gamma'(2)}}\right) = P\left(Y \leq \frac{t^{1/\phi} \lambda}{e^{\Gamma'(2)/\phi}}\right) = \\ &= F_Y\left(\frac{t^{1/\phi} \lambda}{e^{\Gamma'(2)/\phi}}\right) = \left\{ 1 - \exp \left[ - \left( \frac{t^{1/\phi} \lambda / e^{\Gamma'(2)/\phi}}{\lambda / e^{\Gamma'(2)/\phi}} \right)^\phi \right] \right\}^\delta = [1 - \exp(-t)]^\delta. \blacksquare \end{aligned}$$

The  $r$ th ordinary moment of the EW distribution (without restrictions) is given by (Choudhury, 2005)

$$\mu'_k = \lambda_i^k \delta \exp \left[ - \frac{k\Gamma'(2)}{\phi} \right] \Gamma_k(\phi) \left[ 1 + \sum_{i=1}^{\infty} a_i (i+1)^{-(1+\frac{k}{\phi})} \right],$$

where  $a_i = (-1)^i \binom{\delta-1}{i}$ .

The first two moments of the EW distribution are

$$\mu'_1 = E(Y_i) = \lambda_i \delta \exp \left[ - \frac{\Gamma'(2)}{\phi} \right] \Gamma_1(\phi) \left[ 1 + \sum_{i=1}^{\infty} a_i (i+1)^{-(1+\frac{1}{\phi})} \right]$$

and

$$\mu'_2 = E(Y_i^2) = \lambda_i^2 \delta \exp \left[ - \frac{2\Gamma'(2)}{\phi} \right] \Gamma_2(\phi) \left[ 1 + \sum_{i=1}^{\infty} a_i (i+1)^{-(1+\frac{2}{\phi})} \right].$$

So, the expectation and variance of  $Y_i$  are given by

$$E(Y_i) = \lambda_i c_1(\phi, \delta) \quad \text{and} \quad \text{Var}(Y_i) = \lambda_i^2 [h_2(\phi, \delta) - h_1^2(\phi, \delta)] = \lambda_i^2 c_2(\phi, \delta),$$

where

$$c_1(\phi, \delta) = h_1(\phi, \delta) = \delta \exp \left[ - \frac{\Gamma'(2)}{\phi} \right] \Gamma_1(\phi) \left[ 1 + \sum_{i=1}^{\infty} a_i (i+1)^{-(1+\frac{1}{\phi})} \right]$$

and

$$h_2(\phi, \delta) = \delta \exp \left[ - \frac{2\Gamma'(2)}{\phi} \right] \Gamma_2(\phi) \left[ 1 + \sum_{i=1}^{\infty} a_i (i+1)^{-(1+\frac{2}{\phi})} \right].$$

The parameter  $\lambda_i$  is a type of location parameter that also affects the variance function of the EW distribution in the new parametrization so, the variance is a quadratic function of the mean. Plots of the density function of  $Y$  for selected values of  $\delta$  and  $\lambda_i$  are displayed in Figure 2.

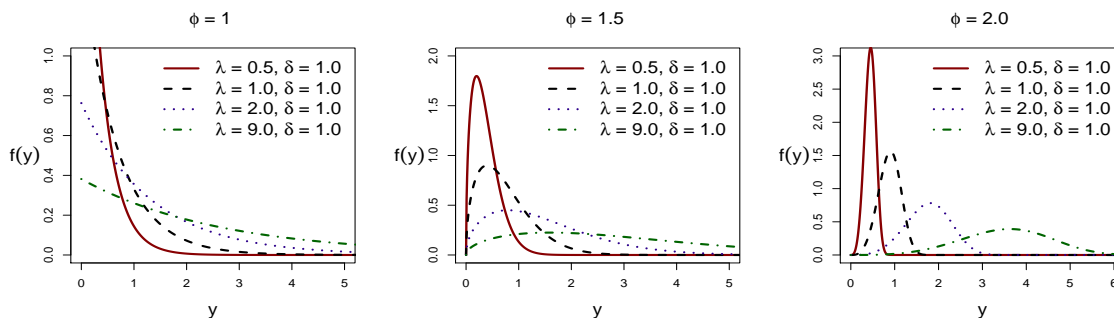


Figure 1. Some densities of the EW distribution

### 3. The New Regression Model

In practice there are many situations where the response variable is influenced by one or more explanatory variables or covariates, and can be related to treatments, intrinsic traits of sample units, exogenous variables, interactions or time-dependent variables, among others. The random component of the GEWLM is defined by the pdf (6) for the response variable  $y_i$ . We accommodate different covariates structures for model (6) using a known function  $g$ , which is twice differentiable and bijective, that links the parameter  $\lambda_i$  to the predictor  $\eta_i$ , called the link function, by

$$g(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta},$$

where  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$  is a set of  $p$  covariates associated with the  $i$ th response variable  $y_i$  for  $i = 1, \dots, n$  and  $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$  is the vector of regression parameters to be estimated. The parameter  $\lambda_i$  can be specified by the inverse of the link function, i.e.  $\lambda_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$  so that  $\mu_i = E(Y_i) = g^{-1}(\eta_i) c_1(\phi, \delta)$ , where it can be seen that part of the mean response for the  $i$ th variable is specified by  $\lambda_i$ , besides influencing the model's variance, because the variance function of the EW distribution is a quadratic function of the mean  $Var(Y_i) = \lambda_i^2 h_2(\phi, \delta) - \mu_i^2$ .

The behavior of  $c_1(\phi, \delta)$  and  $c_2(\phi, \delta)$  are shown in Figure 3 for the values of  $\delta = 0.2, 1, 2$  and  $9$ .

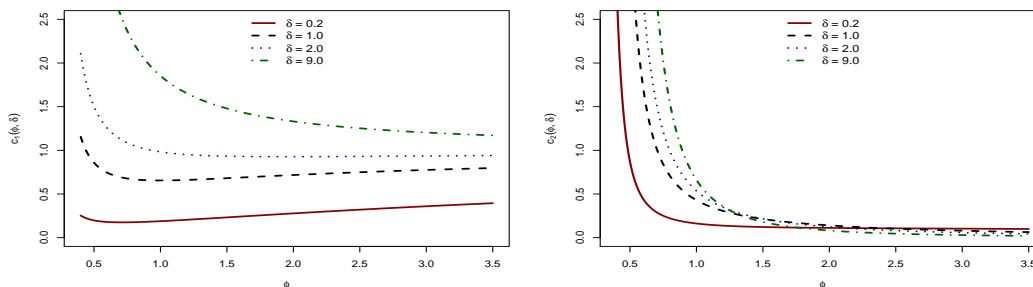


Figure 2. Plots of the functions  $c_1(\phi, \delta)$  and  $c_2(\phi, \delta)$

Figure 3 reveals that as the value of  $\phi$  approaches zero, the expectation and variance of  $Y_i$  increase, while as  $\phi$  increases, both moments decrease. In particular, if  $\delta = 1$ , when  $\phi \rightarrow \infty$ ,  $c_1(\phi, 1) \rightarrow 1$  and  $c_2(\phi, 1) \rightarrow 0$ .

If  $Y_1, \dots, Y_n$  is a sequence of  $n$  independent random variables, such that  $Y_i \sim EW(\lambda_i, \phi, \delta)$  and  $g(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$  is the link function, then the equation that best describes the regression model is given by

$$Y_i = \lambda_i \varepsilon_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) \varepsilon_i, \tag{7}$$

where  $\varepsilon_i \sim EW(1, \phi, \delta)$ . The mean and variance are given, respectively, by

$$\begin{aligned} E(Y_i) &= \lambda_i E(\varepsilon_i) = \lambda_i c_1(\phi, \delta) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) c_1(\phi, \delta) \quad \text{and} \\ \text{Var}(Y_i) &= \lambda_i^2 \text{Var}(\varepsilon_i) = \lambda_i^2 c_2(\phi, \delta) = [g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})]^2 c_2(\phi, \delta). \end{aligned}$$

The systematic component  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$  enters in the regression model in the form of a linear structure for the explanatory variables and the connection between the systematic and random components is established by the inverse link function

$\lambda_i = g^{-1}(\eta_i) > 0$ . The choice of the link function depends strongly on the experiment and the relation between the response variable and the covariates. Some possible choices for the link function are, for example, the identity  $g(\lambda_i) = \lambda_i$ , the reciprocal  $g(\lambda_i) = 1/\lambda_i$  and the logarithmic  $g(\lambda_i) = \log(\lambda_i)$  functions. In this way, the regression is composed of a linear structure in the parameters and a nonlinear function with multiplicative error. Clearly, the GWLM (Prudente & Cordeiro, 2010) is a special case of (7) when  $\delta = 1$ .

#### 4. Estimation, Tests and Simulation

##### 4.1 Maximum Likelihood Estimation

We consider a random sample  $(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)$  of  $n$  independent observations, where  $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$ , for  $i = 1, \dots, n$ , is the vector of covariates associated with the  $i$ th individual and  $y_i$  is the response variable following the EW( $\lambda_i, \phi, \delta$ ) distribution with  $\lambda_i = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$ .

The log-likelihood function for the model parameters  $\boldsymbol{\theta}^T = (\boldsymbol{\beta}^T, \phi, \delta)$  is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \log(\delta) - \log(y_i/\phi) + \log(z_i) - z_i + (\delta - 1) \log[1 - \exp(-z_i)] \right\}, \tag{8}$$

where  $z_i = (y_i/\lambda_i)^\phi e^{r'(2)}$ .

The score vectors for  $\boldsymbol{\beta}$ ,  $\phi$  and  $\delta$  can be expressed as

$$U_{\boldsymbol{\beta}} = \frac{\partial \ell}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{s}, U_{\phi} = \frac{\partial \ell}{\partial \phi} = \frac{n}{\phi} + \boldsymbol{\tau}^T \mathbf{1} \quad \text{and} \quad U_{\delta} = \frac{\partial \ell}{\partial \delta} = \boldsymbol{\kappa}^T \mathbf{1},$$

respectively, where  $\mathbf{W} = \text{diag}[(\lambda'_i/\lambda_i)^2]$  is a diagonal matrix of order  $n$ ,  $\lambda'_i = \partial \lambda_i / \partial \eta_i$ ,  $\lambda''_i = \partial^2 \lambda_i / \partial \eta_i^2$  are the derivatives of the inverse link function,  $\mathbf{1}$  is an  $n \times 1$  vector of ones and the vectors  $\mathbf{s} = (s(z_1), \dots, s(z_n))^T$ ,  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)^T$  and  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_n)^T$  have components given by

$$s(z_i) = (z_i - 1) - (\delta - 1)l_i, \quad \tau_i = -s(z_i) \log(y_i/\lambda_i), \quad \kappa_i = \delta^{-1} + \log[1 - \exp(-z_i)],$$

$$z_i = \left(\frac{y_i}{\lambda_i}\right)^\phi e^{r'(2)}, \quad l_i = \frac{z_i \exp(-z_i)}{1 - \exp(-z_i)}, \quad c_i = \phi \log\left(\frac{y_i}{\lambda_i}\right) = \log\left(\frac{z_i}{e^{r'(2)}}\right) \quad \text{and} \quad f_i = \frac{\lambda''_i \lambda_i}{(\lambda'_i)^2},$$

respectively. The MLEs of the regression coefficients and unknown parameters are the solutions of the nonlinear equations  $U_{\boldsymbol{\beta}} = 0$ ,  $U_{\phi} = 0$  and  $U_{\delta} = 0$ . We use iterative methods to determine the roots. The NLMixed procedure in SAS has been used for maximizing the log-likelihood function  $\ell(\boldsymbol{\theta})$ . Initial values for  $\boldsymbol{\beta}$  and  $\phi$  are taken from the fit of the GWLM with  $\delta = 1$ .

For interval estimation and hypothesis tests on the model parameters, we require the  $(p + 2) \times (p + 2)$  observed ( $\ddot{\mathbf{L}}(\boldsymbol{\theta})$ ) and expected ( $\mathbf{I}(\boldsymbol{\theta})$ ) information matrices given in the Appendix. Under general regularity conditions, we can construct approximate confidence intervals for the individual parameters based on the multivariate normal  $N_{p+2}(0, \mathbf{I}(\hat{\boldsymbol{\theta}})^{-1})$  distribution, where  $\hat{\boldsymbol{\theta}}$  is the MLE of  $\boldsymbol{\theta}$ .

We can compute the maximum values of the unrestricted and restricted log-likelihoods to construct likelihood ratio (LR) statistics for testing some sub-models of the GEWLM. For example, the test of  $H_0 : \delta = 1$  versus  $H_1 : H_0 \text{ is not true}$  is equivalent to comparing the GEWLM and GWLM and the LR statistic reduces to

$$w = 2[\ell(\hat{\boldsymbol{\beta}}, \hat{\phi}, \hat{\delta}) - \ell(\tilde{\boldsymbol{\beta}}, \tilde{\phi}, 1)],$$

where  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\phi}$  and  $\hat{\delta}$  are the MLEs under  $H_1$  and  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\phi}$  are the estimates under  $H_0$ . The test statistic  $w$  has approximately a Chi-square distribution with degree of freedom given by the difference between the numbers of parameters of the two models.

##### Simulation Study: Maximum Likelihood

Next, we perform a Monte Carlo simulation study to evaluate the MLEs of the GEWLM by considering the systematic component  $\eta_i = \beta_0 + \beta_1 x_i, i = 1, \dots, n$ , and the reciprocal ( $g(\lambda_i) = 1/\lambda_i$ ), the identity ( $g(\lambda_i) = \lambda_i$ ) and the logarithmic ( $g(\lambda_i) = \log(\lambda_i)$ ) link functions.

The values of the response variable are simulated from the inverse of the cdf of the EW distribution given by (5), such that  $Y_i = F^{-1}(U_i)$  with  $U_i \sim U(0, 1)$ . Further, the values of the covariates  $x$  are generated by  $x_i = \exp(t_i)$ , where  $t_i \sim N(0, 1)$ , so that  $x$  has strong rightward symmetry.

Therefore, for each combination  $n, \beta_0, \beta_1, \phi, \delta$  and  $g(\lambda_i)$ , we generate 1,000 samples and, for each sample, we obtain the MLEs of the parameters. Further, we calculate the mean, bias and mean squared error (MSE) for each of the 1,000 estimates, as listed in Table 4.1. For this study, we employ four sample sizes ( $n = 75, 150, 300$  and  $750$ ), and consider the reciprocal, identity and logarithmic link functions and true values  $\beta_0 = 1.5, \beta_1 = 2.0, \phi = 0.5, \delta = 0.5$  and  $1.5$ .

Based on the values reported in Table 4.1, we note that, in all cases, as the sample size increases, the the MSEs and biases decrease and the estimates approach the true values. Further, the biases and MSEs are much smaller for the estimates of  $\beta_0$  and  $\beta_1$  using the logarithmic link function, and the convergence to the true parameter values is faster.

Table 1. Mean, bias and MSE based on 1,000 simulations of the GEWLM with  $\beta_0 = 1.5, \beta_1 = 2.0, \phi = 0.5$  and  $\delta = 0.5$  and  $1.5$

$g(\lambda_i)$	n	$\theta$	$\delta = 0.5$			$\theta$	$\delta = 1.5$		
			Average	Bias	MSE		Average	Bias	MSE
$1/\lambda_i$	75	$\beta_0$	3.7136	2.2136	566.7043	$\beta_0$	3.0633	1.5633	52.5531
		$\beta_1$	5.2468	3.2468	1231.4334	$\beta_1$	4.1954	2.1954	158.4900
		$\phi$	0.6288	0.1288	0.1077	$\phi$	0.5806	0.0806	0.0636
		$\delta$	0.5435	0.0435	0.1920	$\delta$	1.8357	0.3357	2.7380
$1/\lambda_i$	150	$\beta_0$	1.7279	0.2279	2.3447	$\beta_0$	2.118	0.6180	7.1914
		$\beta_1$	2.4023	0.4023	3.6811	$\beta_1$	2.7272	0.7272	11.2634
		$\phi$	0.5584	0.0584	0.0304	$\phi$	0.5254	0.0254	0.0167
		$\delta$	0.4999	-0.0001	0.0341	$\delta$	1.6823	0.1823	0.7946
$1/\lambda_i$	300	$\beta_0$	1.6549	0.1549	0.9182	$\beta_0$	1.7592	0.2592	1.0130
		$\beta_1$	2.2022	0.2022	1.6008	$\beta_1$	2.2627	0.2627	1.2709
		$\phi$	0.5250	0.0250	0.0102	$\phi$	0.5073	0.0073	0.0060
		$\delta$	0.5004	0.0004	0.0164	$\delta$	1.5956	0.0956	0.2450
$1/\lambda_i$	750	$\beta_0$	1.5809	0.0809	0.3340	$\beta_0$	1.5818	0.0818	0.2399
		$\beta_1$	2.0757	0.0757	0.3906	$\beta_1$	2.0824	0.0824	0.3840
		$\phi$	0.5065	0.0065	0.0031	$\phi$	0.5044	0.0044	0.0025
		$\delta$	0.5031	0.0031	0.0060	$\delta$	1.5283	0.0283	0.0834
$\lambda_i$	75	$\beta_0$	2.0274	0.5274	5.5805	$\beta_0$	1.9304	0.4304	3.4099
		$\beta_1$	2.3816	0.3816	5.4610	$\beta_1$	2.4132	0.4132	4.5988
		$\phi$	0.6596	0.1596	0.1335	$\phi$	0.5913	0.0913	0.0817
		$\delta$	0.4954	-0.0046	0.0844	$\delta$	1.7533	0.2533	1.5490
$\lambda_i$	150	$\beta_0$	1.6272	0.1272	2.0534	$\beta_0$	1.5747	0.0747	1.1083
		$\beta_1$	2.3445	0.3445	2.9381	$\beta_1$	2.1675	0.1675	1.8551
		$\phi$	0.5681	0.0681	0.0367	$\phi$	0.5225	0.0225	0.0174
		$\delta$	0.4995	-0.0005	0.0406	$\delta$	1.7024	0.2024	0.7972
$\lambda_i$	300	$\beta_0$	1.6210	0.1210	0.9375	$\beta_0$	1.6188	0.1188	0.5708
		$\beta_1$	2.0686	0.0686	1.1237	$\beta_1$	2.0979	0.0979	0.7743
		$\phi$	0.5236	0.0236	0.0102	$\phi$	0.5173	0.0173	0.0072
		$\delta$	0.5038	0.0038	0.0175	$\delta$	1.5510	0.0510	0.2683
$\lambda_i$	750	$\beta_0$	1.5326	0.0326	0.2753	$\beta_0$	1.5656	0.0656	0.2185
		$\beta_1$	2.0224	0.0224	0.4127	$\beta_1$	2.0414	0.0414	0.3016
		$\phi$	0.5067	0.0067	0.0031	$\phi$	0.5085	0.0085	0.0027
		$\delta$	0.5032	0.0032	0.0058	$\delta$	1.5082	0.0082	0.0789
$\log(\lambda_i)$	75	$\beta_0$	1.4416	-0.0584	0.7315	$\beta_0$	1.3060	-0.1940	1.1192

$g(\lambda_i)$	n	$\theta$	$\delta = 0.5$			$\theta$	$\delta = 1.5$		
			Average	Bias	MSE		Average	Bias	MSE
		$\beta_1$	1.9573	-0.0427	0.0455	$\beta_1$	1.9868	-0.0132	0.0138
		$\phi$	0.6329	0.1329	0.1173	$\phi$	0.5668	0.0668	0.0677
		$\delta$	0.5394	0.0394	0.1877	$\delta$	2.0828	0.5828	5.6226
$\log(\lambda_i)$	150	$\beta_0$	1.4700	-0.0300	0.3306	$\beta_0$	1.4362	-0.0638	0.4242
		$\beta_1$	1.9743	-0.0257	0.0168	$\beta_1$	1.9912	-0.0088	0.0054
		$\phi$	0.5507	0.0507	0.0319	$\phi$	0.5295	0.0295	0.0185
		$\delta$	0.5211	0.0211	0.0423	$\delta$	1.6864	0.1864	0.8804
$\log(\lambda_i)$	300	$\beta_0$	1.5056	0.0056	0.1509	$\beta_0$	1.4908	-0.0092	0.1684
		$\beta_1$	1.9842	-0.0158	0.0072	$\beta_1$	1.9940	-0.0060	0.0024
		$\phi$	0.5237	0.0237	0.0104	$\phi$	0.5168	0.0168	0.0074
		$\delta$	0.5070	0.0070	0.0179	$\delta$	1.5545	0.0545	0.2607
$\log(\lambda_i)$	750	$\beta_0$	1.4984	-0.0016	0.0556	$\beta_0$	1.4940	-0.0060	0.0618
		$\beta_1$	1.9938	-0.0062	0.0027	$\beta_1$	1.9980	-0.0020	0.0009
		$\phi$	0.5095	0.0095	0.0032	$\phi$	0.5058	0.0058	0.0025
		$\delta$	0.5019	0.0019	0.0060	$\delta$	1.5211	0.0211	0.0841

### 4.2 Bayesian Estimation

As an alternative analysis, we use the Bayesian method which allows for the incorporation of previous knowledge of the parameters through informative priori density functions.

#### Prior and Posterior Distributions

The normal distribution with mean  $\alpha$  and variance  $\sigma^2$  is denoted by  $N(\alpha, \sigma^2)$ . We assume  $\beta$ ,  $\delta$ , and  $\phi$  are a priori independent, that is,

$$\pi(\theta) = \prod_{i=1}^p \pi(\beta_i)\pi(\delta)\pi(\phi), \tag{9}$$

where  $\beta_j \sim N(0, \sigma_{\beta_j}^2)$ ,  $j = 1, \dots, p$ ,  $\log(\delta) \sim N(0, \sigma_{\delta}^2)$  and  $\log(\phi) \sim N(0, \sigma_{\phi}^2)$ . All the hyper-parameters have been specified to express non-informative priors.

By combining the likelihood function (8) and the prior distribution in (9), the joint posterior distribution for  $\theta$  is obtained as  $\pi(\theta|\mathcal{D}) \propto L(\theta; \mathcal{D}) \prod_{i=1}^p \pi(\beta_i)\pi(\delta)\pi(\phi)$ . This joint posterior density is analytically intractable and we have based our inference on the Markov chain Monte Carlo (MCMC) simulation methods. In particular, the Gibbs sampler algorithm (Gamerman & Lopes, 2006) has proven a powerful alternative. No closed-form is available for any of the full conditional distributions necessary for the implementation of the Gibbs sampler. Thus, we have resorted to the Metropolis–Hastings algorithm. We begin by making a change in the variables to  $\xi = (\log(\delta), \log(\phi), \beta)$ , so that the parameter space is transformed into  $\mathcal{R}^{p+2}$  (necessary for the work with Gaussian densities). Regarding the Jacobian of this transformation, our joint posterior density (or target density) is given by

$$\pi(\xi|\mathcal{D}) \propto L(\xi; \mathcal{D}) \exp \left\{ -\frac{1}{2} \left[ \sum_{j=1}^p \frac{\beta_j^2}{\sigma_{\beta_j}^2} + \frac{\log(\delta)}{\sigma_{\delta}^2} + \frac{\log(\phi)}{\sigma_{\phi}^2} \right] \right\}. \tag{10}$$

To implement the Metropolis-Hastings algorithm, proceed as follows:

- (1) Start with any point  $\xi_{(0)}$  and stage indicator  $j = 0$ ;
- (2) Generate a point  $\xi'$  according to the transitional kernel  $Q(\xi', \xi_j) = N_{p+2}(\xi_j, \tilde{\Sigma})$ , where  $\tilde{\Sigma}$  is the covariance matrix of  $\xi$ , which is the same in any stage;
- (3) Update  $\xi_{(j)}$  to  $\xi_{(j+1)} = \xi'$  with probability  $p_j = \min\{1, \pi(\xi'|\mathcal{D})/\pi(\xi_{(j)}|\mathcal{D})\}$ , or keep  $\theta_{(j)}$  with probability  $1 - p_j$ ;

(4) Repeat steps (2) and (3) by increasing the stage indicator until the process has reached a stationary distribution.

All computations are performed in R software (R Development Core Team [RDCT], 2011).

**Model Comparison Criteria**

A variety of methodologies can be applied for the comparison of several competing models for a given data set and selecting the best one to fit the data. One of the most used approaches is derived from the conditional predictive ordinate (CPO) statistic. For a detailed discussion on the CPO statistic and its applications to model selection, see (Gelfand, Deys & Chang, 1992).

Let  $\mathcal{D}$  denote the full data and  $\mathcal{D}^{(-i)}$  denote the data with the deleted  $i$ -th observation. We denote the posterior density of  $\theta$  given  $\mathcal{D}^{(-i)}$  by  $\pi(\theta|\mathcal{D}^{(-i)})$ ,  $i = 1, \dots, n$ . For the  $i$ -th observation,  $CPO_i$  can be written as

$$CPO_i = \int_{\theta \in \Theta} f(y_i|\theta)\pi(\theta|\mathcal{D}^{(-i)})d\theta = \left\{ \int_{\theta} \frac{\pi(\theta|\mathcal{D})}{f(y_i|\theta)} d\theta \right\}^{-1} \tag{11}$$

The  $CPO_i$  can be interpreted as the height of the marginal density of the time for an event at  $y_i$ . Therefore, high  $CPO_i$  implies a better fit of the model. No closed-form of the  $CPO_i$  is available for the proposed model. However, a Monte Carlo estimate of  $CPO_i$  can be obtained by using a single MCMC sample from the posterior distribution  $\pi(\theta|\mathcal{D})$ . Let  $\theta^{(1)}, \dots, \theta^{(Q)}$  be a sample of size  $Q$  of  $\pi(\theta|\mathcal{D})$  after the burn-in. A Monte Carlo approximation of  $CPO_i$  (Ibrahim, Chen & Sinha, 2001) is given by

$$\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\theta^{(q)})} \right\}^{-1}$$

For model comparisons we use the log pseudo marginal likelihood (LPML) defined by  $LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$ . The higher the  $LPML$  value, the better the fit of the model.

Other criteria, such as the deviance information criterion ( $\widehat{DIC}$ ) proposed by (Spiegelhalter, Best & van der Linde, 2002), the expected Akaike information criterion (EAIC)- (Brooks, 2002), and the expected Bayesian (or Schwarz) information criterion (EBIC)-(Carlin & Louis, 2001) can also be used. They are based on the posterior mean of the deviance, which can be approximated by  $\bar{d} = \sum_{q=1}^Q d(\theta_q)/Q$ , and  $d(\theta) = -2 \sum_{i=1}^n \log [f(y_i|\theta)]$ . The DIC criterion can be estimated using the MCMC output by  $\widehat{DIC} = \bar{d} + \widehat{\rho}_d = 2\bar{d} - \widehat{d}$ , where  $\rho_D$  is the effective number of parameters defined as  $E\{d(\theta)\} - d\{E(\theta)\}$ , where  $d\{E(\theta)\}$  is the deviance evaluated at the posterior mean that can be estimated as

$$\widehat{D} = d\left(\frac{1}{Q} \sum_{q=1}^Q \beta^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \delta^{(q)}, \frac{1}{Q} \sum_{q=1}^Q \phi^{(q)}\right)$$

Similarly, the EAIC and EBIC criteria can be estimated by means of  $\widehat{EAIC} = \bar{d} + 2\#(\theta)$  and  $\widehat{EBIC} = \bar{d} + \#(\theta) \log(n)$ , where  $\#(\theta)$  is the number of model parameters.

**Simulation Study: Bayesian Analysis**

A simulation study conducted to evaluate the parameter estimates for the proposed model through the Bayesian analysis. We consider the systematic component  $\eta_i = \beta_0 + \beta_1 x_i$ ,  $i = 1, \dots, n$ , and the logarithmic ( $g(\lambda_i) = \log(\lambda_i)$ ) link function.

The values of the response variable are simulated from the inverse of the cdf of the EW distribution, such that  $Y_i = F^{-1}(U_i)$  with  $U_i \sim U(0, 1)$ . Further, the values of the covariates  $x$  are generated by  $x_i = \exp(t_i)$ , where  $t_i \sim N(0, 1)$ , so that  $x$  has strong rightward symmetry.

Samples of size  $n = 150$  and  $300$  were generated with logarithmic link functions and true values  $\beta_0 = 2$ ,  $\beta_1 = 1.0$ ,  $\phi = 0.5$ ,  $\delta = 1.5$ . Therefore, each one with 500 Monte Carlo generated data sets. The following values  $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = \sigma_{\delta}^2 = \sigma_{\phi}^2 = 100$  for the prior distributions given in (9) were considered. Note that we assume weak, but informative prior. Because our prior is still informative, the posterior is always proper. In all the work, we consider 40,000 sample burn-in, and use every tenth sample from the 200,000 MCMC posterior samples to reduce the autocorrelations and yield better convergence results, thus obtaining an effective sample of size 20,000 upon which the posterior is based on. We monitor the convergence of the Metropolis-Hasting algorithm using the method proposed by (Geweke, 1992), as well as trace plots.



Further, we calculate the mean, biases, mean squared error (MSE) and 95% Coverage Probability (CP). The figures in Table 4.1 indicate that the bias and MSE decrease when the sample size increases. It is also observed that as the sample size increases the CPs become closer to the nominal value.

Table 2. Mean, bias, MSE and 95% CP for the parameter estimates for fitting the GEWLM model in each dataset

n	Parameter	Mean	Bias	MSE	CP
150	$\delta$	2.0092	0.5092	3.0407	0.916
	$\phi$	0.7096	0.2096	0.6888	0.920
	$\beta_0$	1.8663	-0.1337	1.3791	0.904
	$\beta_1$	1.0058	0.0058	0.0125	0.912
300	$\delta$	1.7135	0.2135	0.7609	0.958
	$\phi$	0.5099	0.0099	0.0137	0.964
	$\beta_0$	1.9000	-0.1000	0.3887	0.954
	$\beta_1$	0.9990	-0.0010	0.0049	0.960

### 5. Sensibility Analysis

There are basically two approaches to detecting observations that seriously influence the results of a statistical analysis. One approach is the influence local approach, and the second approach is Bayesian case influence diagnostics.

#### 5.1 Local Influence

Influence diagnostic is an important step in the analysis of data, since it provides an indication of bad model fit or of influential observations. Since regression models are sensitive to the underlying model assumptions, generally performing a sensitivity analysis is strongly advisable. (Cook, 1986) used this idea to motivate the assessment of influence analysis. He suggested that more confidence can be put in a model which is relatively stable under small modifications. Another approach suggested by (Cook, 1986) is to weight observations instead of removing them. The calculation of local influences can be carried out for model (7). If the likelihood displacement  $LD(\omega) = 2\{l(\hat{\theta}) - l(\hat{\theta}_\omega)\}$  is used, where  $\hat{\theta}_\omega$  denotes the MLE under the perturbed model, then the normal curvature for  $\theta$  at direction  $\mathbf{d}$ ,  $\|\mathbf{d}\| = 1$ , is given by  $C_d(\theta) = 2|\mathbf{d}^T \mathbf{A}^T [\ddot{\mathbf{L}}(\theta)]^{-1} \mathbf{A} \mathbf{d}|$ , where  $\mathbf{A}$  is a  $(p + 2) \times n$  matrix that depends on the perturbation scheme. The elements of this matrix are given by  $\Delta_{vi} = \partial^2 l(\theta|\omega) / \partial \theta_v \partial \omega_i$ ,  $i = 1, 2, \dots, n$  and  $v = 1, 2, \dots, p + 2$ , evaluated at  $\hat{\theta}$  and  $\omega_0$ ;  $\omega_0$  is the no-perturbation vector. For the GEWLM, the elements of  $\ddot{\mathbf{L}}(\theta)$  can be obtained from the authors upon request. We can also calculate the normal curvatures  $C_d(\theta)$  to obtain various index plots, including, for instance, the index plot of  $\mathbf{d}_{max}$ , the eigenvector corresponding to  $C_{d_{max}}$ , the largest eigenvalue of the matrix  $\mathbf{B} = -\mathbf{A}^T [\ddot{\mathbf{L}}(\theta)]^{-1} \mathbf{A}$ , and the index plots of  $C_{\mathbf{d}_i}(\theta)$ , which are together denoted as the total local influence. See, for example, (Lesaffre & Verbeke, 1998), where  $\mathbf{d}_i$  denotes an  $n \times 1$  vector of zeros with one at the  $i$ th position. Thus, the curvature at the direction  $\mathbf{d}_i$  takes the form  $C_i = 2|\mathbf{A}_i^T [\ddot{\mathbf{L}}(\theta)]^{-1} \mathbf{A}_i|$ , where  $\mathbf{A}_i^T$  denotes the  $i$ th row of  $\mathbf{A}$ . It is commonplace to point out cases for which  $C_i \geq 2\bar{C}$ , where  $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$ .

Next, for three perturbation schemes, we calculate the following matrix

$$\mathbf{A} = (\mathbf{A}_{vi})_{(p+2) \times n} = \left[ \frac{\partial^2 l(\theta|\omega)}{\partial \theta_i \partial \omega_v} \right]_{(p+2) \times n}, \quad v = 1, \dots, p + 2 \quad \text{and} \quad i = 1, \dots, n.$$

Previous works on local influence curvatures in regression models for censored data are due to (Escobar & Meeker, 1992), (Ortega, Bolfarine & Paula, 2003), (Ortega, Cancho & Paula, 2009), (Ortega, Cordeiro & Hashimoto, 2011), (Silva, Ortega, Garibay & Barreto, 2008) and (Hashimoto, Ortega, Cancho, & Cordeiro, 2013). We consider model (7) and its log-likelihood function given by (8). We denote the vector of weights by  $\omega = (\omega_1, \dots, \omega_n)^T$ .

- **Case-weight Perturbation**

In this case, the log-likelihood function reduces to

$$\ell(\theta|\omega) = \sum_{i=1}^n \omega_i \left\{ \log(\delta) - \log(y_i/\phi) + \log(z_i) - z_i + (\delta - 1) \log(1 - e^{-z_i}) \right\},$$

where  $z_i = (y_i/\lambda_i)^\phi e^{r(2)}$ ,  $0 \leq \omega_i \leq 1$  and  $\omega_0 = (1, \dots, 1)^T$ . Here,  $\mathbf{A}$  is given by

$$A = \begin{bmatrix} \hat{\phi} X^T \hat{W}^{1/2} \underline{g} \\ (u_1^\phi, \dots, u_n^\phi) \\ (\kappa_1, \dots, \kappa_n) \end{bmatrix},$$

where  $\underline{g} = \text{diag}\{s(\hat{z}_1), \dots, s(\hat{z}_n)\}$ ,  $u_i^\phi = \hat{\phi}^{-1}[1 - s(\hat{z}_i) \log(\hat{z}_i/e^{\Gamma(2)})]$  and  $\kappa_i = \delta^{-1} + \log(1 - e^{-z_i})$ .

• **Response Perturbation**

Now, we consider that each  $y_i$  is perturbed as  $y_{i\omega} = y_i + \omega_i S_y$ , where  $S_y$  is a scale factor that may be estimated by the standard deviation of the observed response  $y$  and  $\omega_i \in \mathfrak{R}$ . The perturbed log-likelihood function can be expressed as

$$\ell(\theta|\omega) = \sum_{i=1}^n \left\{ \log(\delta) - \log(y_{i\omega}/\phi) + \log(z_{i\omega}) - z_{i\omega} + (\delta - 1) \log[1 - e^{-z_{i\omega}}] \right\},$$

where  $z_{i\omega} = (y_{i\omega}/\hat{\lambda}_i)^{\hat{\phi}} e^{\Gamma(2)}$  and  $\omega_0 = (0, \dots, 0)^T$ . The matrix  $A$  is given by

$$A = \hat{\phi} S_y \begin{bmatrix} \hat{\phi} X^T \hat{W}^{1/2} \underline{g} \\ -(u_{1\omega}^\phi/\hat{\phi} y_{1\omega}, \dots, u_{n\omega}^\phi/\hat{\phi} y_{n\omega}) \\ (l_{1\omega}/y_{1\omega}, \dots, l_{n\omega}/y_{n\omega}) \end{bmatrix},$$

where  $\underline{g} = \text{diag}\{n_{1\omega}/y_{1\omega}, \dots, n_{n\omega}/y_{n\omega}\}$ ,  $u_{i\omega}^\phi = s(z_{i\omega}) + n_{i\omega} \log(z_{i\omega}/e^{\Gamma(2)})$ ,  $n_{i\omega} = z_{i\omega} - (\delta - 1)l_{i\omega}(1 - e^{-z_{i\omega}})$ ,  $s(z_{i\omega}) = (z_{i\omega} - 1) - (\delta - 1)l_{i\omega}$  and  $l_{i\omega} = z_{i\omega} e^{-z_{i\omega}} / (1 - e^{-z_{i\omega}})$ .

• **Explanatory Variable Perturbation**

Consider now an additive perturbation on a particular continuous explanatory variable, say  $x_j$ , by setting  $x_{ij\omega} = x_{ij} + \omega_i S_{x_j}$ , where  $S_{x_j}$  is a scale factor and  $\omega_i \in \mathfrak{R}$ . The perturbed log-likelihood function has the form

$$\ell(\theta|\omega) = \sum_{i=1}^n \left\{ \log(\delta) - \log\left(\frac{y_i}{\phi}\right) + \log(z_{i\omega}) - z_{i\omega} + (\delta - 1) \log[1 - e^{-z_{i\omega}}] \right\},$$

where  $z_{i\omega} = (y_i/\hat{\lambda}_{i\omega})^{\hat{\phi}} e^{\Gamma(2)}$ ,  $\hat{\lambda}_{i\omega} = g^{-1}(x_i^T \hat{\beta} + S_{x_j} \hat{\beta}_j \omega_i)$  and  $\omega_0 = (0, \dots, 0)^T$ . The matrix  $A$  is given by

$$A_{p+2 \times n} = (A_{p \times n}^\beta, A_{1 \times n}^\phi, A_{1 \times n}^\delta)^T,$$

where

$$\begin{aligned} A_{ij}^\beta &= \begin{cases} \hat{\phi} \hat{\beta}_t S_{x_j} x_{ij} (\lambda'_{i\omega}/\lambda_{i\omega})^2 [(f_{i\omega} - 1)s(z_{i\omega}) - \hat{\phi} n_{i\omega}], & j \neq t, \\ \hat{\phi} \hat{\beta}_t S_{x_j} x_{ij} (\lambda'_{i\omega}/\lambda_{i\omega})^2 [(f_{i\omega} - 1)s(z_{i\omega}) - \hat{\phi} n_{i\omega}] + \hat{\phi} S_{x_j} (\lambda'_{i\omega}/\lambda_{i\omega}) s(z_{i\omega}), & j = t, \end{cases} \\ A_i^\phi &= \hat{\phi} \hat{\beta}_t S_{x_j} (\lambda'_{i\omega}/\lambda_{i\omega}) [\hat{\phi}^{-1} s(z_{i\omega}) + n_{i\omega} c_{i\omega}], \\ A_i^\delta &= -\hat{\phi} \hat{\beta}_t S_{x_j} (\lambda'_{i\omega}/\lambda_{i\omega}) l_{i\omega}. \end{aligned}$$

Note that the equations,  $\lambda'_{i\omega}$ ,  $s(z_{i\omega})$ ,  $l_{i\omega}$ ,  $n_{i\omega}$ ,  $c_{i\omega}$  and  $f_{i\omega}$  are defined in Section 3, but applied to  $z_{i\omega}$  and  $\lambda_{i\omega}$ .

In order to determine whether the  $i$ th observation is possibly influential, (Poon & Poon, 1999) proposed classifying the  $i$ th observation as possible influential if  $M(0)_i = \frac{Cd_i(\theta)}{tr(2T)}$  is greater than  $\overline{M}(0) + c^* S M(0)$ , where  $\overline{M}(0) = 1/q$ ,  $S M(0)$  is the sample standard error  $\{M(0)_k, k = 1, \dots, q\}$  and  $c^*$  is a any constant selected according to the real application.

5.2 Bayesian Case Influence Diagnostics

(Cook, 1986) suggested that more confidence should be put in a model relatively stable under small modifications. The best known perturbation schemes are based on case-deletion (Cook & Weisberg, 1982) whose effects are studied of completely removing cases from the analysis. This reasoning will be the basis for our Bayesian global influence methodology for the determining of the influential observation in the analysis.

This reasoning will form the basis of our Bayesian global influence methodology and, in doing so, it will be possible to determine which subjects might influence the analysis. In this work, we use the Bayesian case-deletion influence

diagnostic measures for the joint posterior distribution based on the  $\psi$ -divergence ((Peng & Dey, 1995) and (Weiss, 1996)).

Let  $D_\psi(P, P_{(-i)})$  denote the  $\psi$ -divergence between  $P$  and  $P_{(-i)}$ , in which  $P$  denotes the posterior distribution of  $\vartheta$  for the full data, and  $P_{(-i)}$  denotes the posterior distribution of  $\vartheta$  without the  $i$ th case. Specifically,

$$D_\psi(P, P_{(-i)}) = \int \psi \left( \frac{\pi(\theta|\mathcal{D}^{(-i)})}{\pi(\theta|\mathcal{D})} \right) \pi(\theta|\mathcal{D}) d\theta, \tag{12}$$

where  $\psi$  is a convex function with  $\psi(1) = 0$ . Several choices concerning the  $\psi$  are given by (Dey & Birmiwal, 1994). For example,  $\psi(z) = -\log(z)$  defines the Kullback-Leibler (K-L) divergence,  $\psi(z) = 0.5|z - 1|$  defines the  $L_1$  norm (or variational distance) and  $\psi(z) = (z - 1) \log(z)$  gives the  $J$ -distance (or the symmetric version of K-L divergence).

Let  $\theta^{(1)}, \dots, \theta^{(Q)}$  be a sample of  $\pi(\theta|\mathcal{D})$ ,  $D_\psi(P, P_{(-i)})$  the calculated numerically by

$$\widehat{D}_\psi(P, P_{(-i)}) = \frac{1}{Q} \sum_{q=1}^Q \psi \left( \frac{\widehat{CPO}_i}{f(y_i|\theta^{(q)})} \right), \tag{13}$$

where  $\widehat{CPO}_i = \left\{ \frac{1}{Q} \sum_{q=1}^Q \frac{1}{f(y_i|\theta^{(q)})} \right\}^{-1}$  is the numerical approximation of the conditional predictive ordinate statistic of the  $i$ -th observation (Ibrahim et al., 2001).

Note that  $D_\psi(P, P_{(-i)})$  can be interpreted as the  $\psi$ -divergence of the effect of deleting the  $i$ -th case from the full data on the joint posterior distribution of  $\theta$ . As pointed out by (Peng & Dey, 1995) and (Weiss, 1996), it may be difficult for a practitioner to judge the cutoff point of the divergence measure so as to determine whether a small subset of observations is influential or not. Therefore, we will use the proposal by (Peng & Dey, 1995) and (Weiss, 1996) by considering a biased coin, which has success probability  $p$ . Then, the  $\psi$ -divergence between the biased and unbiased coins is

$$D_\psi(f_0, f_1) = \int \psi \left( \frac{f_0(x)}{f_1(x)} \right) f_1(x) dx, \tag{14}$$

where  $f_0(x) = p^x(1 - p)^{1-x}$  and  $f_1(x) = 0.5$ ,  $x = 0, 1$ . If  $D_\psi(f_0, f_1) = d_\psi(p)$ , it can be easily checked that  $d_\psi$  satisfies the following equation

$$d_\psi(p) = \frac{\psi(2p) + \psi(2(1 - p))}{2}. \tag{15}$$

It is not difficult to see for the divergence measures considered that  $d_\psi$  increases as  $p$  moves away from 0.5. In addition,  $d_\psi(p)$  is symmetric at  $p = 0.5$  and  $d_\psi$ , achieves its minimum at  $p = 0.5$ . At this point,  $d_\psi(0.5) = 0$ , and  $f_0 = f_1$ . Therefore, if we consider  $p > 0.90$  (or  $p \leq 0.10$ ) as a strong bias in a coin, then  $d_{K-L}(0.90) = 0.51$ ,  $d_J(0.90) = 0.88$  and  $d_{L_1}(0.90) = 0.4$ . Thus, if we use the  $J$ -distance, an observation which  $d_J > 0.88$  can be considered influential. Similarly, using the Kullback-Leibler divergence and the  $L_1$  norm, we can consider an influential observation when  $d_{K-L} > 0.51$  and  $d_{L_1} > 0.4$ , respectively.

### 5.3 Influence of Outlying Observations

One of our main goals is to show the need for robust models to deal with the presence of outliers in the data. We consider simulated datasets with one, two and three generated perturbed cases to examine the performance of the proposed diagnostics measures. A sample of size 300 is generated by GEWLM considering the logarithmic link function and true values  $\beta_0 = 1, \beta_1 = 0.5, \phi = 2, \delta = 1.5$ . In the simulated data,  $y_i$  ranged from 0.4535 to 744.8722 with median = 4.1141, mean = 13.4078 and standard deviation = 56.6212. We selected the cases 25, 150 and 225 for perturbation. To create an influential observation in the dataset, we choose one to three select the cases and perturbed the response variable as follows:  $\tilde{y}_i = y_i + 5S_y, i = 25, 150$  and 225, where  $S_y$  is the standard deviations of the  $y_i$ 's. Here, we consider eight setups in the study. Setup A: original dataset, without outliers; Setup B: data with outlier 25; Setup C: data with outlier 150; Setup D: data with outliers 225; Setup E: data with outliers 25 and 150; Setup F: data with outlier 25 and 225; Setup G: data with outliers 150 and 225; and Setup H: data with outliers 25, 150 and 225. The MCMC computations are made similar to those in the last subsection and further to monitor the convergence of the Gibbs samples we use the Geweke's convergence diagnostic proposed by (Geweke, 1992).

Table 5.3 shows the posterior inferences about the parameters that, except for the parameter  $\beta_1$ , are sensitive to the perturbation of the selected case(s).

Table 5.3 shows the Monte Carlo estimates of the DIC, EAIC, EBIC and LPML criteria for each perturbed version of the original data set. We can observe that, as expected, the original simulated data (Setup A) has the best fit.

Table 3. Mean and standard deviation (SD) for the parameter estimates for fitting the GEWLM in each dataset

Dataset names	Perturbed case	Parameters			
		$\delta$	$\phi$	$\beta_0$	$\beta_1$
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
A	None	1.6337 (0.5221)	1.8532 (0.2945)	0.9896 (0.1141)	0.4901 (0.0143)
B	25	27.4966 (5.4909)	0.4909 (0.0276)	-1.0455 (0.2077)	0.4798 (0.0194)
C	150	20.6355 (4.8647)	0.5605 (0.0391)	-0.6994 (0.2218)	0.4899 (0.0186)
D	225	26.6348 (5.3982)	0.4992 (0.0290)	-1.0024 (0.2100)	0.4817 (0.0191)
E	{25, 150}	30.9432 (5.7190)	0.4504 (0.0235)	-1.2664 (0.2087)	0.4819 (0.0202)
F	{25, 225}	33.2884 (5.7556)	0.4310 (0.0211)	-1.3888 (0.2032)	0.4763 (0.0213)
G	{150, 225}	30.5419 (5.7452)	0.4547 (0.0247)	-1.2439 (0.2112)	0.4832 (0.0198)
H	{25, 150, 225}	35.1426 (6.2281)	0.4078 (0.0222)	-1.5410 (0.2254)	0.4796 (0.0224)

Table 4. Bayesian criteria

Dataset names	Criterion			
	EAIC	EBIC	DIC	LPML
A	1,376.4892	1,391.3043	1,371.7413	-686.1304
B	1,502.2160	1,517.0310	1,497.8440	-755.9290
C	1,477.5376	1,492.3528	1,473.2525	-742.6665
D	1,497.5984	1,512.4136	1,493.3490	-753.0548
E	1,542.6485	1,557.4637	1,538.3057	-772.7122
F	1,555.7940	1,570.6092	1,551.7796	-780.6969
G	1,539.2131	1,554.0282	1,535.1550	-772.1113
H	1,592.2849	1,607.1000	1,589.8817	-796.7907

For each simulated data set, we now consider the sample from the posterior distributions of the parameters of the EW model to calculate the three  $\psi$ -divergence measures ( $d_{KL}, d_J, d_{L_1}$ ) described in Section 5.2. The results in Table 5.3 show, before perturbation (Setup A), the selected cases are not influential according to all  $\psi$ -divergence measures. However, after perturbation (Setup, B-H), the measures increase, which indicates that the perturbed cases are influential. Thus, we clearly see that all  $\psi$ -divergence measures perform well to identifying influential case(s),

Table 5.  $\psi$ -divergence measures for the simulated data fitting the WE model

Dataset Names	Case Number	$\psi$ -divergence measures		
		$d_{K-L}$	$d_J$	$d_{L_1}$
A	25	0.0050	0.0100	0.0401
	150	0.0028	0.0057	0.0300
	225	0.0155	0.0312	0.0702
B	25	7.6356	13.4340	0.9523
C	150	6.5771	11.5979	0.9426
D	225	7.0510	12.0335	0.9352
E	25	3.2524	5.8476	0.7939
	150	1.1009	2.1237	0.5379
F	25	3.0605	6.0520	0.8067
	225	2.7802	5.5341	0.7852
G	150	1.4175	2.8145	0.6120
	225	4.0247	7.1761	0.8459
H	25	1.5909	3.0451	0.6263
	150	0.6040	1.1972	0.4204
	225	1.5199	2.9159	0.6156

Figures 5.3, 5.3, 5.3 and 5.3 show the three  $\psi$ -divergence measures for datasets A, B, F and H, respectively. All measures identified influential case(s) and provided larger  $\psi$ -divergence measures in comparison to the other cases.

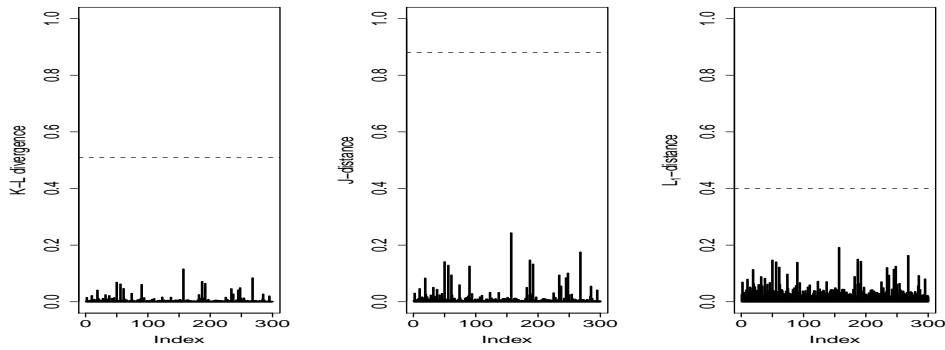


Figure 3.  $\psi$ -divergence measures from dataset A

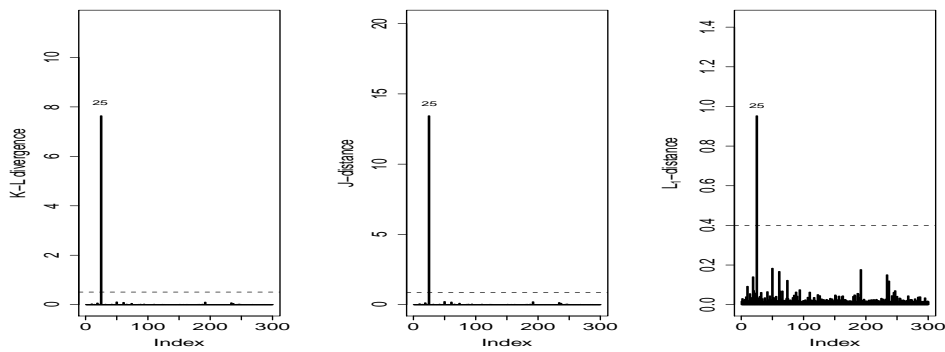


Figure 4.  $\psi$ -divergence measures from dataset B

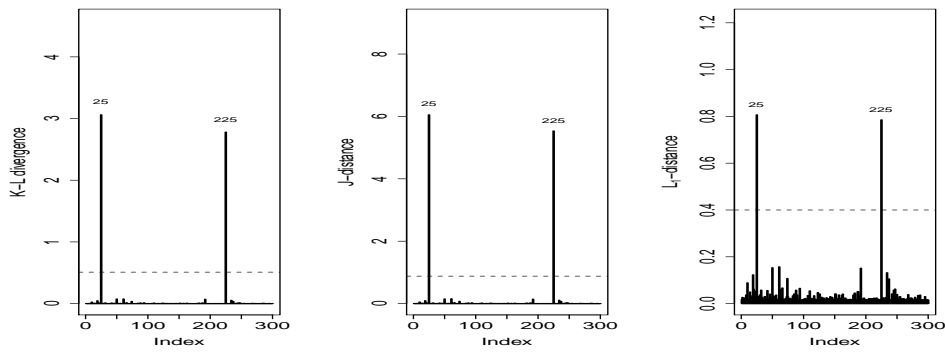


Figure 5.  $\psi$ -divergence measures from dataset F

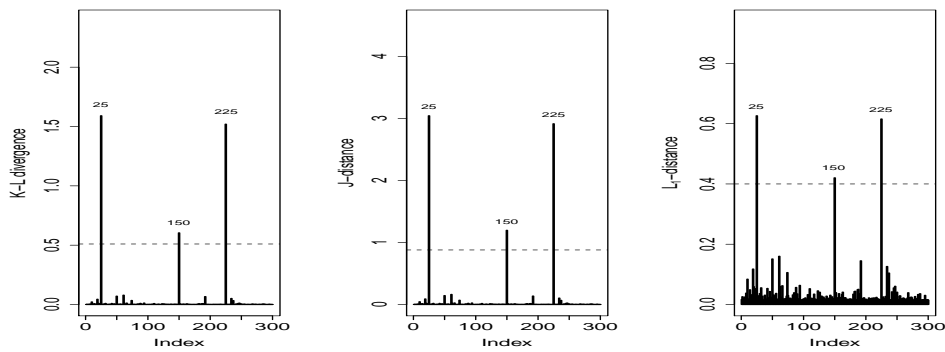


Figure 6.  $\psi$ -divergence measures from dataset H

## 6. Residual Analysis

When attempting to adjust a model to a dataset, the validation of this fit must be analyzed by a specific statistic with the purpose, with the purpose of measuring the goodness-of-fit. Once the model is chosen and fitted, the analysis of the residuals is an efficient way to check the model's adequacy. The residuals also serve for other purposes, such as to detect the presence of aberrant points (outliers), identify the relevance of an additional factor omitted from the model and verify if there are indications of serious deviance from the distribution assumed for the random error. Further, since the residuals are used to identify discrepancies between the fitted model and the dataset, it is convenient to try to define residuals that take into account the contribution of each observation to the goodness-of-fit measure used.

In summary, the residuals allow measuring the model's fit for each observation and enable studying whether the differences between the observed and fitted values are due to chance or to a systematic behavior that can be modeled. The methods proposed by (Dunn & Smyth, 1996) can be used to obtain quantile residuals.

The quantile residual for the GEWLM is given by

$$t_i = \Phi^{-1}\{[1 - \exp(-z_i)]^\delta\},$$

where  $\Phi(\cdot)$  is the cumulative standard normal distribution.

(Atkinson, 1985) suggested the construction of envelopes to enable better interpretation of the normal plot of probabilities of the residuals. These envelopes are simulated confidence bands that contain the residuals, such that if the model is well fitted, the majority of points will be within these bands and randomly distributed. The construction of the confidence bands follows these steps

- Fit the proposed model and calculate the residuals  $t_i$ 's;
- Simulate  $k$  samples of the response variable using the fitted model;
- Fit the model to each sample and calculate the residuals  $t_{ij}$ ,  $j = 1, 2, \dots, k$  and  $i = 1, 2, \dots, n$ ;
- Arrange each group of  $n$  residuals in rising order to obtain  $t_{(i)j}$  for  $j = 1, 2, \dots, k$  and  $i = 1, 2, \dots, n$ ;
- For each  $i$ , obtain the mean, minimum and maximum  $t_{(i)j}$ , namely

$$t_{(i)M} = \sum_{j=1}^k \frac{t_{(i)j}}{k}, \quad t_{(i)I} = \min\{t_{(i)j} : 1 \leq j \leq k\} \quad \text{and} \quad t_{(i)S} = \max\{t_{(i)j} : 1 \leq j \leq k\};$$

- Include the means, minimum and maximum together with the values of  $t_i$  against the expected percentiles of the standard normal distribution.

The minimum and maximum values of  $t_i$  form the envelope. If the model under study is correct, the observed values should be inside the bands and distributed randomly.

## 7. Application: Insurance Payouts for Personal Accidents

The dataset comes from a study of the size of the indemnities paid for personal accidents by Australian insurers, in the period from July 1989 to June 1999, published by (Jong & Heller, 2008). The aim is to relate the variables legal representation (1 if the insured is represented by a lawyer and 0 otherwise) and claim settlement time, representing the percentage of cases whose settlement time was faster than the average for all claims of the same type, with the payouts denominated in Australian dollars.

The indemnity amounts paid by insurers are typically concentrated in a low range, with many fewer high payouts, so that the distribution has rightward asymmetry of the data. There are various distributions with rightward asymmetry and positive support that can be used to model these data. Therefore, application of the hrf is an important tool to model these data, because often distributions with totally different failure function behavior have similar densities.

For this analysis, we consider the last 18 months of the study (January 1998 to June 1999) and insureds with legal representation, giving a total of 542 observations, with settlement time as the covariate. Initially, we conduct an exploratory analysis of the data, whose results are presented in Table 7. The histogram of this dataset is displayed in Figure 7a. The results in Table 7 and Figure 7 show symmetry to the right of the data, indicating adjustment by an asymmetric distribution is most appropriate.

In many applications there is qualitative information about the hazard shape, which can help with selecting a particular GEWLM. In this context, a device called the total time on test (TTT) plot (Aarset, 1987) is useful. The TTT plot is obtained by plotting  $G(r/n) = [(\sum_{i=1}^r T_{i:n}) + (n - r)T_{r:n}]/(\sum_{i=1}^n T_{i:n})$ , where  $r = 1, \dots, n$  and  $T_{i:n}$  for  $i = 1, \dots, n$  are the order statistics of the sample, against  $r/n$ . It is a straight diagonal for constant hazards leading to an exponential model. It is convex for decreasing hazards and concave for increasing hazards leading to a single-Weibull model. It is first convex and then concave if the hazard is bathtub-shaped leading to a GEWLM. It is first concave and then convex if the hazard is bimodal-shaped leading to a GEWLM. For multimodal hazards, the TTT plot contains several concave and convex regions.

The TTT-plot for this dataset is displayed in Figure 7b, where it can be seen that the curve is initially concave and then becomes convex around the reference diagonal line, indicating that the general shape of the hrf is unimodal, so the GEWLM is appropriate for the data.

It is interesting to note that if the TTT-plot curve had initially been convex and then turned concave, the hrf would have been U-shaped, so the Weibull distribution would not be suitable to fit the data, since it does not model data with bathtub hrf, unlike the EW distribution.

Table 6. Descriptive statistics of the dataset of personal accident insurance payouts

Median	Mean	Std. Dev.	Coeff. of variation	Asymmetry	Kurtosis	Min	Max	n
6765.25	8996.41	8790.80	97.71 %	5.59	54.32	109	116586.72	542

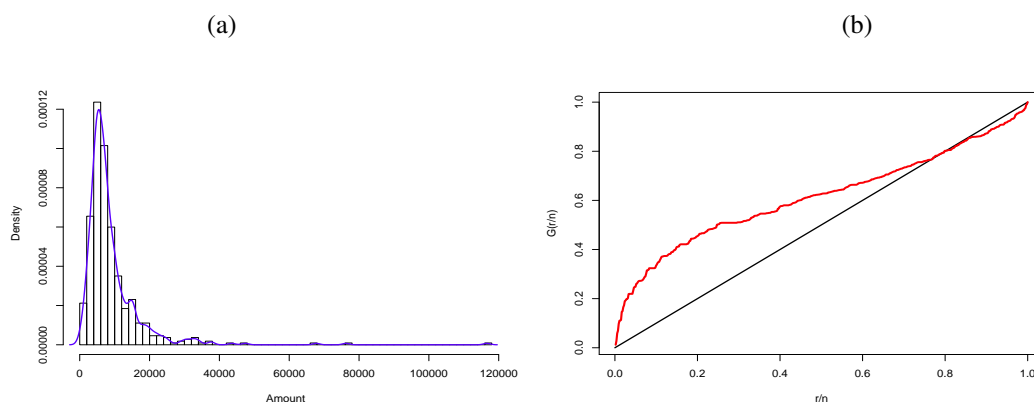


Figure 7. Personal accident insurance payouts (a) Histogram (b) TTT-plot

### 7.1 Maximum Estimation

The regression model, considering the GEWLM, is given by

$$y_i = g^{-1}(\beta_0 + \beta_1 x_i) \epsilon_i, \quad i = 1, \dots, 542,$$

where  $y_i$  is the amount paid from the  $i$ th observation,  $x_i$  is the settlement time associated with the  $i$ th observation,  $\epsilon_i$  is the random error having the  $EW(1, \phi, \delta)$  distribution and  $g^{-1}(\cdot)$  is the inverse link function.

In Figure 7, we provide the dispersion plots of the settlement time covariate: (a) versus the response variable (amount of the indemnity); and (b) versus the natural logarithm of the response variable. It can be seen that the logarithm of the indemnity amount grows linearly with settlement time and the variability is stabilized. Therefore, we adopt the logarithmic link function  $\log(\lambda_i) = \eta_i$  to fit the GEWLM, and the regression model is given by

$$y_i = \exp(\beta_0 + \beta_1 x_i) \epsilon_i .$$

Table 7 gives the Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC) and Bayesian Information Criterion (BIC), the estimates of the parameters, the standard errors (SEs) and estimated levels ( $p$ -values) for the GEWLM and GWLM regression models. It can be noted that the GEWLM again gives the lower values for the AIC, CAIC and BIC statistics than the GWLM. The regression parameters are significant for both models, so they should remain in the study.

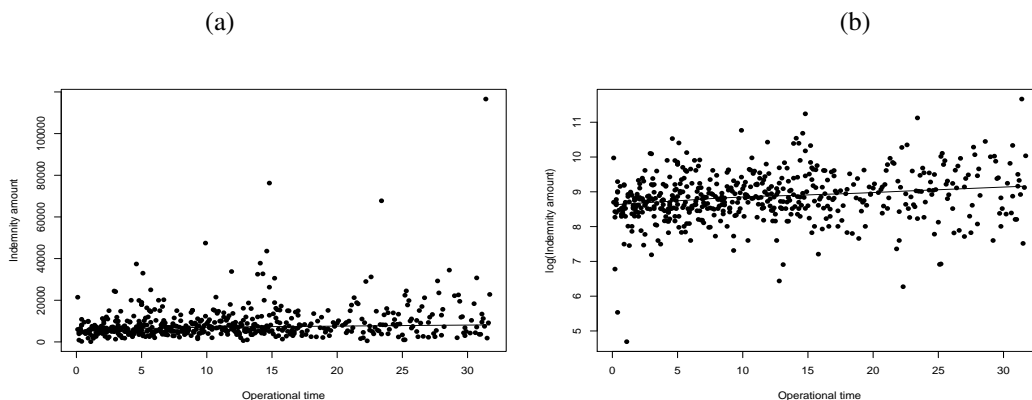


Figure 8. Dispersion plots of the covariate settlement time ( $x_i$ ). (a) Versus the response variable indemnity amount ( $y_i$ ). (b) Versus  $\log(y_i)$ .

Next, we calculate the odds ratio statistic to test the hypotheses  $H_0 : \delta = 1$  versus  $H_1 : \delta \neq 1$  to compare the GEWLM and GWLM, namely  $w = 2[-5365.0 - (-5413.0)] = 96$  ( $p$ -value =  $< 0.0001$ ). Then, according to the hypothesis test,  $H_0$  should not be accepted. Therefore, the best model to fit the data is the GEWLM.

Table 7. MLE of the parameters for the GWLM and GEWLM regression models

Model	$\delta$	$\phi$	$\beta_0$	$\beta_1$
GEWLM	6.6226	0.6859	8.0552	0.01769
SE	1.6332	0.0623	0.2047	0.0036
$p$ -value	-	-	$< 0.0001$	$< 0.0001$
AIC = 10738	CAIC = 10739	BIC = 10736		
Model	-	$\phi$	$\beta_0$	$\beta_1$
GWLM	-	1.4022	9.1651	0.0273
SE	-	0.0412	0.0527	0.0036
$p$ -value	-	-	$< 0.0001$	$< 0.0001$
AIC = 10832	CAIC = 10832	BIC = 10845		

Further, we perform an analysis of local influence for the current data using the GEWLM. The plots in Figure 7 indicate the possible influential observations in the GEWLM regression. The plots of local influence ( $M(0)$  against the index of the observations) considered three perturbation schemes: likelihood (Figure 7a), response variable (Figure 7b) and the covariable settlement time (Figure 7c). The observations 28, 369 and 539 appear in the three perturbation schemes as possible influential points, so we remove them from the dataset, individually and together, and re-estimated the model parameters to study the influence of these observations on the estimates of the parameters. The observation 28 is the lowest indemnity payouts for personal accident damages by the Australian insurers within the sample studied ( $y_{28} = 109$ ), with settlement time  $x_{28} = 1.1$  and the observations 369 and 539 are the highest ( $y_{369} = 76255.76$ ,  $x_{369} = 14.8$ ) and ( $y_{539} = 116586.72$ ,  $x_{539} = 31.4$ ).

Table 7 gives the relative changes (in percents) of the estimates of the parameters after removal of the observations 28, 369, 539, (28 and 369), (28 and 539), (369 and 539) and (28, 369 and 539), defined by  $RC_{\theta_j} = (|\hat{\theta}_j - \hat{\theta}_{j(i)}|/\hat{\theta}_j) \times 100$ , where  $\hat{\theta}_{j(i)}$  is the MLE of  $\theta_j$  after removing the  $i$ th observation. It can be seen that the significance of the regression parameters  $\beta_0$  and  $\beta_1$  did not change, indicating the model’s robustness. Hence, there is no change in the inference after removing the possible influential observations (cases 28, 369, 539, (28 and 369), (28 and 539), (369 and 539) and (28, 369 and 539)).

7.2 Bayesian Analysis

We have checked the sensitivity analysis for the variance component parameters for various choices of prior parameters. Here, we consider four priors in the study. Prior 1:  $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = \sigma_{\delta}^2 = \sigma_{\phi}^2 = 50$ ; Prior 2:  $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = \sigma_{\delta}^2 = \sigma_{\phi}^2 = 100$ ; Prior 3:  $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = \sigma_{\delta}^2 = \sigma_{\phi}^2 = 1,000$  and Prior 4:  $\sigma_{\beta_0}^2 = \sigma_{\beta_1}^2 = \sigma_{\delta}^2 = \sigma_{\phi}^2 = 10,000$ .

The posterior summaries of the parameters do not present remarkable difference and not impair the results in Table 7.

In an overall sense, confronting the GEWLM and GWLM, since in Table 7 the estimate of  $\delta$  is different from 1,  $1 \notin (4.1627, 10.3994)$ , we have indications favoring the GEWLM model. Also, we present in Table 7 the values of the DIC,



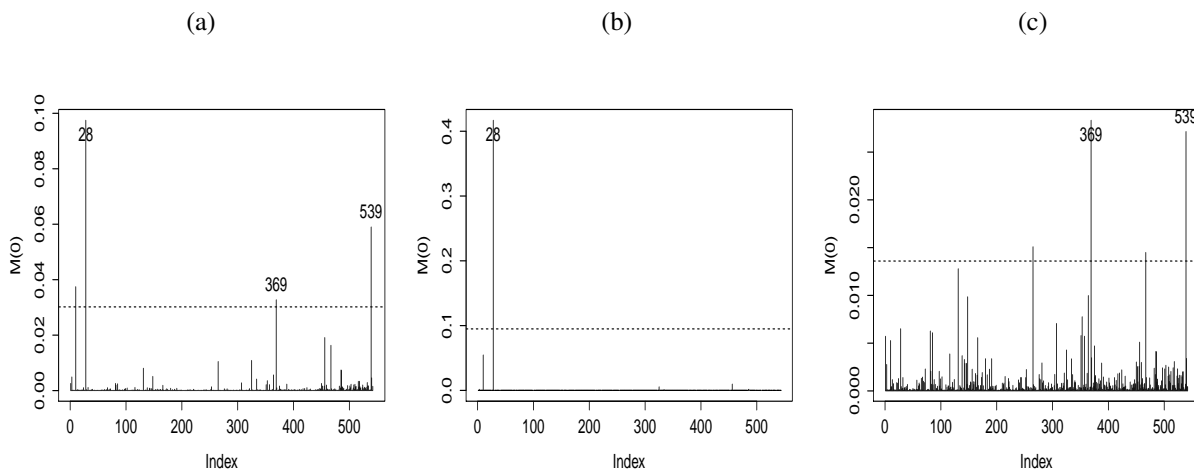


Figure 9. Index plot of  $\mathbf{d}_{max}$  for  $\theta$ . (a) Case-weights perturbation. (b) Response perturbation. (c) Explanatory variable perturbation

Table 8. Relative changes [RC in %], standard errors and  $p$ -values in parentheses when the observations 28, 369, 539, (28 and 369), (28 and 539), (369 and 539) and (28, 369 and 539) are deleted, respectively.

Dropping	$\hat{\delta}$	$\hat{\phi}$	$\hat{\beta}_0$	$\hat{\beta}_1$
Full	[-]	[-]	[-]	[-]
	1.6332	0.0623	0.2047	0.0036
	(-)	(-)	(< 0.0001)	(< 0.0001)
28	[37.9445]	[8.7768]	[3.0142]	[9.1577]
	2.6763	0.06255	0.2574	0.0036
	(-)	(-)	(< 0.0001)	(< 0.0001)
369	[11.2203]	[5.7734]	[1.2265]	[1.1306]
	1.4259	0.0667	0.1920	0.0035
	(-)	(-)	(< 0.0001)	(< 0.0001)
539	[12.1504]	[6.4587]	[1.5977]	[8.7620]
	1.3996	0.0667	0.1920	0.0035
	(-)	(-)	(< 0.0001)	(< 0.0001)
28 and 369	[20.1818]	[3.2512]	[1.4798]	[7.5549]
	2.2717	0.0667	0.2391	0.0035
	(-)	(-)	(< 0.0001)	(< 0.0001)
28 and 539	[18.4303]	[2.4931]	[1.0813]	[16.1673]
	2.2189	0.0669	0.2379	0.0035
	(-)	(-)	(< 0.0001)	(< 0.0001)
369 and 539	[23.9008]	[14.0545]	[2.9968]	[8.0837]
	1.1946	0.0729	0.1792	0.0035
	(-)	(-)	(< 0.0001)	(< 0.0001)
28, 369 and 539	[0.6070]	[4.6946]	[0.6257]	[14.9802]
	1.8385	0.0728	0.2195	0.0034
	(-)	(-)	(< 0.0001)	(< 0.0001)

EAIC, EBIC and LPML criteria to compare these models. The GEWLM stands out as the best one.

We compute the  $\psi$ -divergence measures in (12) described in Section 5.2. Table 7 gives the results for the four observations that have higher values. Cases 28 and 359 are possible influential observations in the posterior distribution. The observation 28 is the lowest indemnity payouts for personal accident damages by the Australian insurers within the sample studied ( $y_{28} = 109$ ), with settlement time  $x_{28} = 1.1$  and the observation 539 is the highest ( $y_{539} = 116586.72$ ,  $x_{539} = 31.4$ ).

From now on, we assume Prior 2. Figures 7, 7, 7 and 7 reveal the approximate marginal posterior densities of the model parameters and trace plot considering the 10,000 observations generated, and Figure 7 shows the index plot of the three  $\psi$ -divergence measures.

Figure 7 shows the index plots of  $\psi$ -divergence measures for the insurance payouts for personal accidents data when we dropped observations 28 and 359. For all divergence measures, we can see that after dropped the two observations have

Table 9. Real data. Posterior summaries of the parameters for the GEWLM and GWLM

Parameter	GEWLM			GWLM			
	Mean	SD	CI (95%)	Mean	SD	CI (95%)	
Prior 1	$\delta$	6.6338	1.6052	(4.1627, 10.3994)	-	-	-
	$\phi$	0.6936	0.0605	(0.5810, 0.8152)	1.3985	0.0412	(1.3187, 1.4797)
	$\beta_{01}$	8.0662	0.1957	(7.6559, 8.4145)	9.1668	0.0536	(9.0636, 9.2735)
	$\beta_{11}$	0.0178	0.0036	(0.0107, 0.0249)	0.0273	0.0036	(0.0202, 0.0344)
Prior 2	$\delta$	6.8091	1.7166	(4.1717, 10.8384)	-	-	-
	$\phi$	0.6882	0.0620	(0.5713, 0.8156)	1.3985	0.0412	(1.3182, 1.4807)
	$\beta_{01}$	8.0463	0.2052	(7.6086, 8.4146)	9.1664	0.0533	(9.0648, 9.2737)
	$\beta_{11}$	0.0178	0.0036	(0.0108, 0.0249)	0.0273	0.0036	(0.0202, 0.0345)
Prior 3	$\delta$	6.9665	1.8359	(4.2487, 11.3956)	-	-	-
	$\phi$	0.6835	0.0628	(0.5635, 0.8104)	1.3983	0.0408	(1.3189, 1.4789)
	$\beta_{01}$	8.0299	0.2133	(7.5671, 8.4041)	9.1675	0.0529	(9.0669, 9.2723)
	$\beta_{11}$	0.0177	0.0036	(0.0106, 0.0248)	0.0272	0.0036	(0.0202, 0.0343)
Prior 4	$\delta$	7.0313	1.8745	(4.2484, 11.4799)	-	-	-
	$\phi$	0.6816	0.0632	(0.5619, 0.8098)	1.3980	0.0409	(1.3179, 1.4805)
	$\beta_{01}$	8.0223	0.2164	(7.5569, 8.3995)	9.1670	0.0538	(9.0640, 9.2757)
	$\beta_{11}$	0.0177	0.0036	(0.0105, 0.0249)	0.0273	0.0037	(0.0200, 0.0345)

Table 10. Real data. Bayesian criteria

Model	Criterion				
	EAIC	EBIC	DIC	LPML	
Prior 1	GEWLM	10,742.440	10,759.620	10,738.190	-5,370.800
	GWLM	10,835.162	10,848.048	10,832.194	-5,418.818
Prior 2	GEWLM	10,742.486	10,759.667	10,738.244	-5,370.903
	GWLM	10,835.130	10,848.016	10,832.132	-5,418.637
Prior 3	GEWLM	10,742.531	10,759.712	10,738.262	-5,371.059
	GWLM	10,835.110	10,847.990	10,832.080	-5,418.720
Prior 4	GEWLM	10,742.539	10,759.721	10,738.252	-5,370.985
	GWLM	10,835.137	10,848.023	10,832.141	-5,418.749

Table 11.  $\psi$ -divergence measures for the real data fitting the GEWLM

Case Number	$\psi$ -divergence measures			
	$d_{K-L}$	$d_J$	$d_{L_1}$	
Prior 1	10	0.2577	0.5386	0.2842
	28	0.7266	1.5899	0.4687
	369	0.2566	0.5543	0.2824
	539	0.4666	0.9917	0.3830
Prior 2	10	0.2688	0.5583	0.2918
	28	0.7482	1.5887	0.4759
	369	0.2567	0.5462	0.2835
	539	0.4785	1.0395	0.3847
Prior 3	10	0.2929	0.6127	0.3031
	28	0.8646	1.9346	0.5083
	369	0.2523	0.5327	0.2816
	539	0.4668	1.0065	0.3809
Prior 4	10	0.2821	0.5820	0.2981
	28	0.7976	1.6793	0.4882
	369	0.2548	0.5432	0.2838
	539	0.4658	1.0026	0.3796

not detected more influential points.

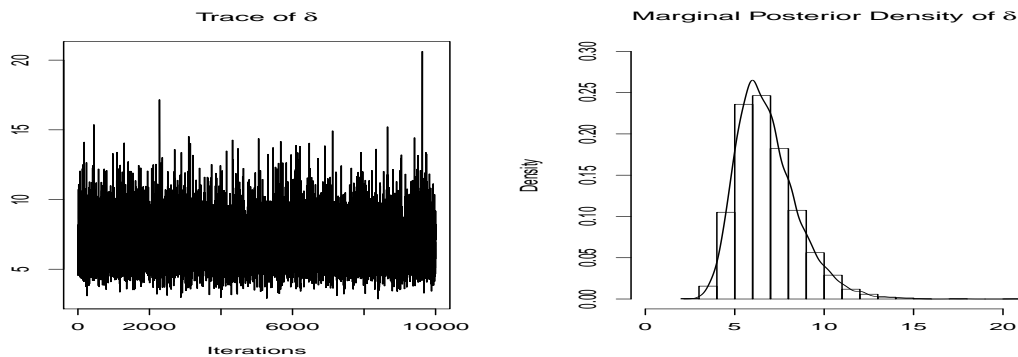


Figure 10. Trace plot and approximate marginal posterior density for the parameter  $\delta$

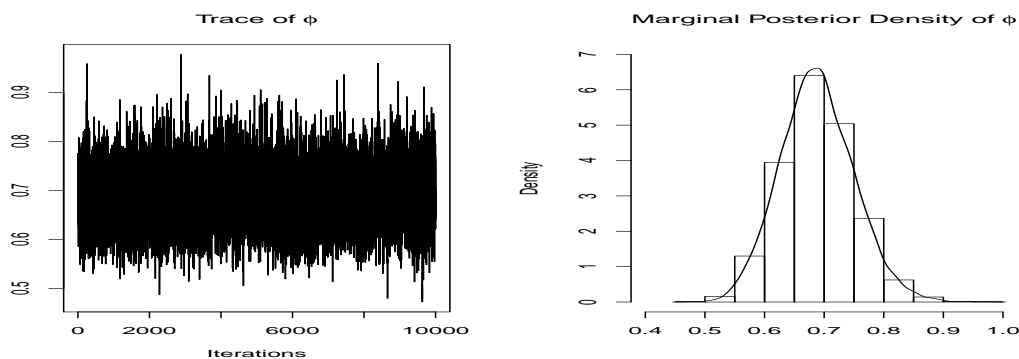


Figure 11. Trace plot and approximate marginal posterior density for the parameter  $\phi$ .

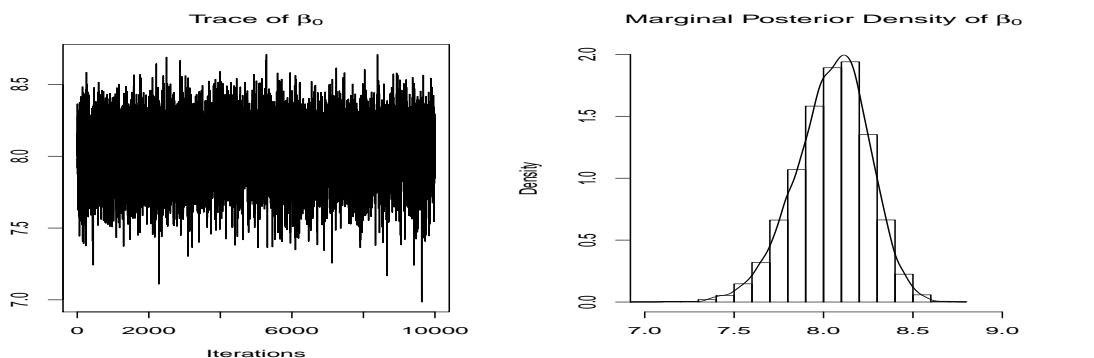


Figure 12. Trace plot and approximate marginal posterior density for the parameter  $\beta_0$

Table 7 gives the relative change (RCs) of each estimate of the parameters after removal of the observations 28, 539, and (28 and 539), and the 95% credible intervals for the parameters. The RC (in percentage) of each estimated parameter is defined by  $RC_{\theta_j} = |(\hat{\theta}_j - \hat{\theta}_{j(I)})/\hat{\theta}_j| \times 100\%$ , where  $\hat{\theta}_{j(I)}$  denotes the posterior mean of  $\theta_j$ , with  $j = 1, \dots, 4$ , after set I of observations has been removed. It can be seen a small RC for the parameter  $\beta_0$  and that the significance of the regression parameters  $\beta_0$  and  $\beta_1$  did not change, indicating the model's robustness. Furthermore, a better fit was achieved when we dropped observation 28, in comparison with the fitting when we removed the observation 539 according to, according to all criteria.

Thus, Table 7 and Table 7 reveal the parameter estimates, standard errors and significance of the parameters for the MLEs and Bayesian estimates, respectively. By examining the figures in this table, we conclude that the estimates by the two methods are very similar. We show that the GEWLM seems to be more appropriate for fitting the data set than the GWLM.

We can also see in the diagnostics analysis results similar in both approaches.

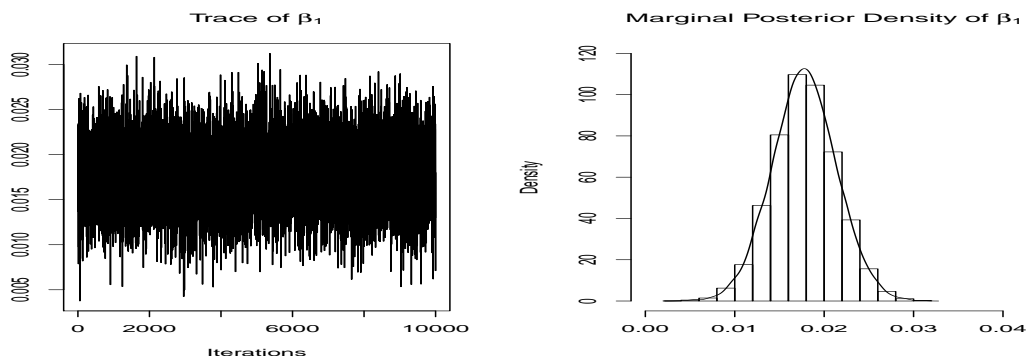


Figure 13. Trace plot and approximate marginal posterior density for the parameter  $\beta_1$

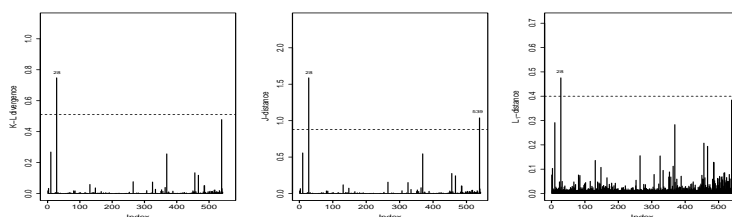


Figure 14.  $\psi$ -divergence measures from full real dataset

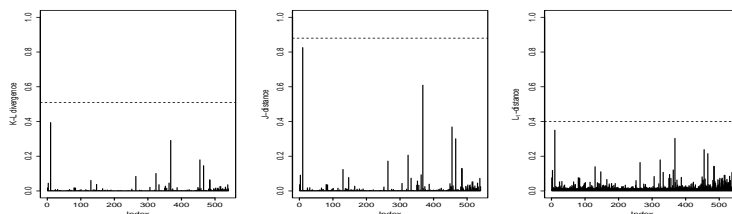


Figure 15.  $\psi$ -divergence measures from real dataset when when dropped observations 28 and 539

Also, we present residual analysis. Figure 7 displays the Normal probability plot for the quantile residuals  $t_i$ 's and the simulated envelope, which are used to check the regression model's goodness-of-fit. The Normal probability plot is constructed using the complete dataset, because the possible influential points (10 and 28) do not modify the significance of the estimates. The broken red lines in Figure 7a are determined by observation, so that the points farthest from  $t_i = 0$  are highlighted. Figure 7b presents the simulated envelope, where the blue lines represent  $t_{(i)J}$ ,  $t_{(i)M}$  and  $t_{(i)S}$ . The confidence band is defined by the interval  $[t_{(i)J}, t_{(i)M}]$ , shown in Figure 7b, because it must contain the values  $t_i$  (black dots) for the set. Note that all points are inside the envelope, indicating the good fit of the GEWLM. The final model is thus given by

$$y_i = \exp(8.0552 + 0.01769x) = 3150.13 \times 1.02^{x_i}.$$

Therefore, an increase of 2% in the indemnity paid for personal accidents implies an increase in the settlement time of one unit, or 1%.

Table 12. RCs (in %) and the corresponding 95% credible interval (CI) for fitting the WE model to the data when dropped the influential cases

Parameter	Dropped Observation					
	28		539		{28, 539}	
	RC	CI (95%)	RC	CI (95%)	RC	CI (95%)
$\delta$	33.6288	(5.2704, 15.3334)	11.9081	(3.7309, 9.5729)	16.4957	(4.6816, 13.0944)
$\phi$	7.5214	(0.5248, 0.7590)	6.3285	(0.6074, 0.8661)	1.6903	(0.5573, 0.8077)
$\beta_0$	2.6203	(7.3133, 8.2615)	1.5869	(7.7516, 8.5185)	0.8646	(7.4998, 8.3779)
$\beta_1$	8.5665	(0.0092, 0.0233)	9.2832	(0.0093, 0.0232)	16.1857	(0.0081, 0.0218)
EAIC	10,708.190		10,701.520		10,667.745	
EBIC	10,725.370		10,718.693		10,684.911	
DIC	10,703.800		10,697.318		10,663.368	
LPML	-5,353.220		-5,350.178		-5,332.747	

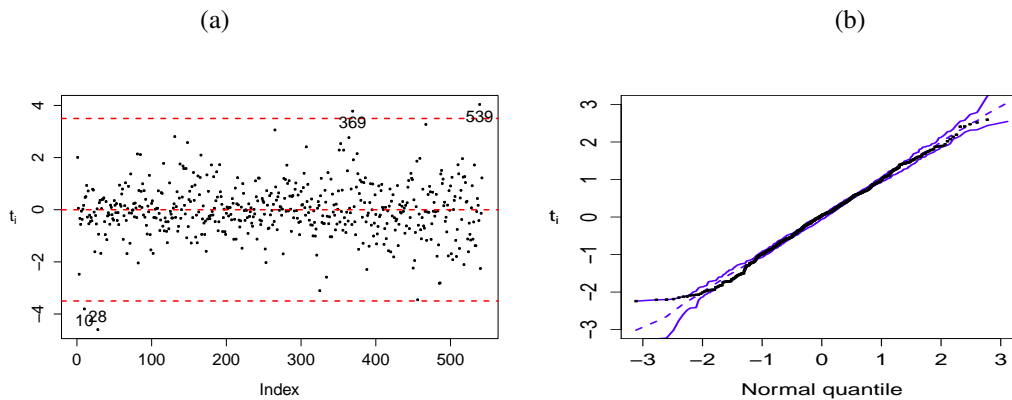


Figure 16. (a) Quantile residuals ( $t_i$ ) versus the indices of the observations for the GEWLM. (b) Normal probability plot for the quantile residuals with simulated envelope for the GEWLM

**8. Conclusions**

The generalized exponentiated Weibull linear model (GEWLM) brings a new approach to the exponentiated Weibull (EW) distribution in regression models. It is effective in modeling a part of the mean response in positive asymmetric data, including the explanatory variables by means of a linear function in the parameters called the linear predictor ( $\eta = X\beta$ ) in a nonlinear model using a link function  $g(\lambda) = \eta$ , which connects the symmetric part of the model to the random part (in the same way of the generalized linear models). The estimation of the parameters are determined by two different methods: maximum likelihood and a Bayesian approach. We use the local influence and Bayesian case influence diagnostics in order to study the sensitivity of the GEWLM. This model is more flexible than the generalized linear Weibull model (GLWM) proposed by (Prudente & Cordeiro, 2010), since it can model monotone and non-monotone hazard rate functions and has the GWLM as a sub-model. In the application, the GEWLM provides a better fit than the GWLM. The new model is robust to small perturbations in the data and in the model, and all the residual values remain within for the proposed residuals remain within the confidence bands of the simulated envelope constructed, indicating the adequacy of the fit.

**References**

Aarset, M. V. (1987). How to identify bathtub rate. *Transactions on Reliability*, 36, 106-108. <https://doi.org/10.1109/TR.1987.5222310>

Ahmad, N., Islam, A., & Salam, A. (2006). Analysis of optimal accelerated life test plans for periodic inspection: the case of exponentiated Weibull failure model. *International Journal of Quality and Reliability Management*, 23, 1019-1046. <https://doi.org/10.1108/02656710610688194>

Atkinson, A. C. (1985). *Plots, transformations and regressions*. Oxford Statistical Sciences Series, Oxford: Clarendon

Press.

- Barghout, M. (2009). An exponentiated Weibull software reliability model. *Advances and Applications in Statistics*, 13, 111-130.
- Carlin, B. P., & Louis, T. A. (2001). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Boca Raton: Chapman and Hall.
- Choudhury, A. (2005). A simple derivation of moments of the exponentiated Weibull Distribution. *Metrika*, 62, 17-22. <https://doi.org/10.1007/s001840400351>
- Cox, D. R., & Reid, N. (1987) Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society, Series B*, 49, 1-39.
- Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society, Series B*, 48, 133-169.
- Cook, R. D., & Weisberg, S. (1982). *Residuals and Influence in Regression*. Boca Raton: Chapman and Hall.
- Collet, A. (2003). *Modelling Survival Data in Medical Research*. London: Chapman and Hall.
- Dey, D., & Birmiwil, L. (1994). Robust Bayesian analysis using divergence measures. *Statistics & Probability Letters*, 20, 287-294. [https://doi.org/10.1016/0167-7152\(94\)90016-7](https://doi.org/10.1016/0167-7152(94)90016-7)
- Dunn, K. P., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236-244. <https://doi.org/10.1080/10618600.1996.10474708>
- Escobar, L. A., & Meeker, W. Q. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, 48, 507-528. <https://doi.org/10.2307/2532306>
- Gamerman, D., & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd. Boca Raton: Chapman and Hall.
- Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. *Bayesian Statistics*, 4, 147-167.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*. eds: J. Bernardo et al. *Oxford University Press*, 1, 169-188.
- Hashimoto, E. M., Ortega, E. M. M., ..., & Cordeiro, G. M. (2013). On estimation and diagnostics analysis in log-generalized gamma regression model for interval-censored data. *Statistics*, 47, 379-398. <https://doi.org/10.1080/02331888.2011.605888>
- Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian Survival Analysis*. Springer, New York. <https://doi.org/10.1007/978-1-4757-3447-8>
- Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*. Cambridge University Press: Cambridge, England. <https://doi.org/10.1017/CBO9780511755408>
- Lesaffre, E., & Verbeke, G. (1998). Local influence in linear mixed models, *Biometrics*, 54, 570-582. <https://doi.org/10.2307/3109764>
- Mudholkar, G. S., & Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Communication in Statistics - Theory and Methods*, 25, 3059-3083.
- Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-ratedata. *IEEE Transactions on Reliability*, 42, 299-302. <https://doi.org/10.1109/24.229504>
- Nadarajah, S., Cordeiro, G. M., & Ortega, E. M. M. (2013). The exponentiated Weibull distribution: a survey. *Statistical Papers*, 54 839-877. <https://doi.org/10.1007/s00362-012-0466-x>
- Ortega, E. M. M., Bolfarine, H., & Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, 42, 165-186.
- Ortega, E. M. M., Cancho, V. G., & Paula G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*, 15, 79-106.
- Ortega, E. M. M., Cordeiro, G. M., & Hashimoto, E. M. (2011). A log-linear regression model for the Beta-Weibull distribution. *Communications in Statistics - Simulation and Computation*, 40, 1206-1235.
- Poon, W., & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal*

*Statistical Society, Series B, 61, 51-61.*

Peng, F. & Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics - La Revue Canadienne de Statistique, 23*, 199-213.

Prudente, A. A., & Cordeiro, G. M. (2010). Generalized Weibull linear models. *Communications in Statistics - Theory and Methods, 39*, 3739-3755.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Silva, G. O., Ortega, E. M. M., ..., & Barreto, M. L. (2008). Log-Burr XII regression models with censored Data *Computational Statistics and Data Analysis, 52*, 3820-3842.

Spiegelhalter, D. J., Best, N.G., ..., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B, 64*, 583-639. <https://doi.org/10.1111/1467-9868.00353>

Surles, J. G., & D'ambrosio, D. M. (2004). A Burr type X chain-of-links model for carbon fibers. *Journal of Composite Materials, 38*, 1337-1343. <https://doi.org/10.1177/0021998304042735>

Vanegas, L. H., Rondón, L. M., & Cordeiro, G. M. (2013). Diagnostic tolls in generalized Weibull linear regression models. *Journal of Statistical Computation and Simulation, 83*(12), 2315-2338. <https://doi.org/10.1080/00949655.2012.690764>

Wang, M., & Rennolls, K. (2005). Tree diameter distribution modelling: introducing the logit-logistic distribution. *Can J For Res, 35*, 1305-1313. <https://doi.org/10.1139/x05-057>

Weiss, R. (1996). An approach to Bayesian sensitivity analysis. *Journal of the Royal Statistical Society Series B, 739-750.*

Zhang, T., Xie, M., ..., & Ng, S. H. (2005). Reliability and modeling of systems integrated with firmware and hardware. *International Journal of Reliability Quality and Safety Engineering, 12*, 227-239.

**Appendix**

**Observed Information Matrix  $\ddot{L}(\theta)$**

Here, we derive the formulas to obtain the second-order partial derivatives of the log-likelihood function. After some algebraic manipulations, we obtain

$$\ddot{L}(\theta) = \begin{bmatrix} \ddot{L}_{\beta\beta} & \ddot{L}_{\beta\phi} & \ddot{L}_{\beta\delta} \\ \ddot{L}_{\phi\beta} & \ddot{L}_{\phi\phi} & \ddot{L}_{\phi\delta} \\ \ddot{L}_{\delta\beta} & \ddot{L}_{\delta\phi} & \ddot{L}_{\delta\delta} \end{bmatrix} = \begin{bmatrix} -\phi^2 \mathbf{X}^T \mathbf{W} [d] \mathbf{X} & \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{U}_\phi & -\phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{L} \\ - & -\frac{n}{\phi^2} - \frac{1}{\phi^2} \mathbf{M}^T \mathbf{1} & -\frac{1}{\phi} \mathbf{B}^T \mathbf{1} \\ - & - & -\frac{n}{\delta^2} \end{bmatrix},$$

where

$$\mathbf{B} = (b_1, \dots, b_n)^T, \quad \mathbf{d} = \text{diag}[d_1, \dots, d_n], \quad \mathbf{U}_\phi = (u_1^\phi, \dots, u_n^\phi)^T, \\ \mathbf{M} = (m_1, \dots, m_n)^T, \quad \mathbf{L} = (l_1, \dots, l_n)^T,$$

$$b_i = c_i l_i, \quad c_i = \phi \log\left(\frac{y_i}{\lambda_i}\right) = \log\left(\frac{z_i}{e^{f'(z)}}\right), \quad f_i = \frac{\lambda_i'' \lambda_i}{(\lambda_i')^2},$$

$$d_i = \frac{1}{\phi} \{ \phi z_i - s(z_i)(f_i - 1) - \phi(\delta - 1)l_i[1 - \exp(z_i)l_i] \},$$

$$m_i = n_i c_i^2, \quad n_i = z_i - (\delta - 1)l_i [1 - \exp(z_i)l_i],$$

$$u_i^\phi = s(z_i) + n_i c_i.$$

**Expected Information Matrix  $\mathbf{I}(\theta)$**

The elements of the expected information matrix are given by

$$\mathbf{I}(\theta) = \begin{bmatrix} \mathbf{I}_{\beta\beta} & \mathbf{I}_{\beta\phi} & \mathbf{I}_{\beta\delta} \\ \mathbf{I}_{\phi\beta} & \mathbf{I}_{\phi\phi} & \mathbf{I}_{\phi\delta} \\ \mathbf{I}_{\delta\beta} & \mathbf{I}_{\delta\phi} & \mathbf{I}_{\delta\delta} \end{bmatrix} = \begin{bmatrix} \phi^2 \mathbf{X}^T \mathbf{W} \mathbf{V} \mathbf{X} & \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{O} & \phi \delta t_1 \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{1} \\ - & -\frac{n}{\phi^2} - \frac{1}{\phi^2} \mathbf{G}^T \mathbf{1} & -\frac{n\delta}{\phi^2} t_2 \\ - & - & \frac{n}{\delta^2} \end{bmatrix},$$

where

$$\mathbf{V} = \text{diag}(v_1, \dots, v_n)^T, \mathbf{O} = (o_1, \dots, o_n)^T, \mathbf{G} = (g_1, \dots, g_n)^T,$$

$$T(d, b, a, e) = \int_0^\infty x^{d-1} e^{-bx} (1 - e^{-x})^{a-1} [\log(x)]^e dx,$$

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad \psi(p) = \frac{\partial \log[\Gamma(p)]}{\partial p}, \quad \psi'(x) = \frac{\partial^2 \log[\Gamma(x)]}{\partial x^2},$$

$$t_1 = T(2, 2, \delta - 1, 0), \quad t_2 = T(2, 2, \delta - 1, 1) - \psi(2)T(2, 2, \delta - 1, 0),$$

$$v_i = \delta(\delta - 1)T(3, b + 1, \delta - 2, 0) + 1,$$

$$o_i = -\psi(2) + \frac{1}{B(\delta, 1)} \{T(2, 1, \delta, 1) + (\delta - 1)[T(3, 0, \delta - 1, 1) - T(2, 2, \delta - 1, 1) - \psi(2)T(3, 2, \delta - 2, 0)]\},$$

$$g_i = \frac{1 - (\delta - 1)}{B(\delta, 1)} \{T(2, 1, \delta, 2) - 2\psi(2)T(2, 1, \delta, 1) + B(\delta, 1)[\psi(2)]^2\} + \frac{\delta - 1}{B(\delta, 1)} \{T(3, 2, \delta - 2, 2) - 2\psi(2)T(3, 2, \delta - 2, 1) + [\psi(2)]^2 T(3, 2, \delta - 2, 0)\}$$

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license. (<http://creativecommons.org/licenses/by/4.0/>).