International Journal of Development Mathematics Vol 2 Issue 3 Page No. 248 - 258

INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: https://ijdm.org.ng/index.php/Journals

On Robust Regression: A Model with Weibull Error Term

Isaac E. Gongsin[a*], Samaila J. Yaga[a] and Na'awurti W. Nyandaiti[b]

[a]*Department of Statistics, University of Maiduguri, Nigeria*

[b]*School of Health Information Management, University of Maiduguri Teaching Hospital*

ARTICLE INFO

Article history:
Received 08 April 2025
Received in revised form 20 August 2025
Accepted 10 September 2025

Keywords:
Generalized method, least absolute deviation, likelihood ratio test, Newton-Raphson, wind speed.

MSC 2020 Subject classification:
62J99

ABSTRACT

A regression model based on the Weibull error distribution has been developed to address the challenges of modeling skewed response variables commonly encountered in environmental and life sciences data. By introducing a location parameter to the standard Weibull distribution, the model accommodates asymmetry and heavy tails that linear normal-based regression often fails to capture. Parameter estimation was carried out using the Newton–Raphson algorithm, implemented in the R program WeiReg. Model comparison based on the Akaike Information Criterion (AIC) demonstrates that the Weibull regression provides a substantially better fit than the linear regression model when applied to wind speed data from Christmas Island, Australia. These findings highlight the relevance of flexible error structures in regression modeling and underscore the practical value of the Weibull approach for analyzing environmental data.

1. Introduction

The linear regression model is widely applied due to its simplicity and intuitive appeal in parameter estimation and inference. The model was first pioneered by Francis Galton (Gujarati *et al*., 2012), and subsequently developed by Karl Pearson and Sir R. A. Fisher (Pollock, 2014). Fisher specifically made the more remarkable contribution toward the development of modern day regression analysis through experimentation (Aldrich, 2005). Its application is based on the assumption of normality in the response and predictor variables, parameters and the error term. Independence among values of the error term (which has zero mean and constant finite variance), and non-correlation among predictor variables in the case of multiple regression models are also assumed (Wooldridge, 2013).

However, data generation processes in most practical situations are not amenable to the normal distribution. For instance, live time data often produce skewed observations restricted to positive values; thus the normal distribution, which spans all real values is often inadequate for such data. Deviations from normality have been studied since the end of the nineteenth century; see for instance, Baissa and Rainey (2020), Huber (1981) and Tiku *et al*. (1986). Regression models and analysis procedures with non-normal

* Corresponding author. Tel.: +23408085000240
E-mail address: gongsin2013@unimaid.edu.ng (.Isaac E Gongsin)
https://doi.org/10.62054/ijdm/0203.17

errors are found in Ganguly (2014), Lindsey (1997) and Manning *et al*. (2003).

Normality assumptions have been attractive because of the ease of computation and inference about parameter estimates. But the advent of computers has made complex computations much easier in a minimum time frame. This paper considers the regression model analysis in the case where the response variable, and consequently the error term, assume the Weibull distribution; that is, the weaker assumption that the error term is Weibull distributed is made. This model is appropriate for the study of causal relationship in natural and environmental phenomena; for instance, solar and wind energy for hybrid power modeling, length of drought and quantity of harvest per hectare of land particularly in rain-dependent agrarian society like Nigeria, among others.

2. Model Construction

Let $(x_i, y_i), i = 1, 2, \dots, n$ be independent random observations of the random variables (X, Y), where X has the density function f and the conditional distribution of Y given X at x is Weibull (Gongsin and Saporu, 2020), given by

$$f_{Y/X}(y/X = x) = \frac{\lambda}{x}\left(\frac{y-\mu}{x}\right)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}}, \qquad y > \mu, \qquad x, \lambda, \mu > 0 \tag{1}$$

We desire to find the conditional expectation of Y given X at x as follows

$$E(Y/X = x) = \int_{\mu}^{\infty} y f_{Y/X}(y/X = x)\, dy$$

$$= \int_{\mu}^{\infty} y \frac{\lambda}{x}\left(\frac{y-\mu}{x}\right)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

$$= \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} (y - \mu + \mu)\,(y-\mu)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

$$= \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} (y-\mu)\,(y-\mu)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy + \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} \mu\,(y-\mu)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

$$= A + B$$

$$A = \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} (y-\mu)\,(y-\mu)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

$$= \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} (y-\mu)^{\lambda} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

Let $z = \left(\frac{y-\mu}{x}\right)^{\lambda}$, then $y-\mu = xz^{\frac{1}{\lambda}}$ and $dy = \frac{x}{\lambda} z^{\frac{1}{\lambda}-1} dz$; $y = \mu \implies z = 0$ and $y = \infty \implies z = \infty$

$$A = \frac{\lambda}{x^{\lambda}} \int_{0}^{\infty} \left(xz^{\frac{1}{\lambda}}\right)^{\lambda} e^{-z} \frac{x}{\lambda} z^{\frac{1}{\lambda}-1} dz$$

$$= x \int_{0}^{\infty} z^{\frac{1}{\lambda}+1-1} e^{-z}\, dz$$

$$= x\Gamma_{\left(1+\frac{1}{\lambda}\right)}$$

$$B = \frac{\lambda}{x^{\lambda}} \int_{\mu}^{\infty} \mu\,(y-\mu)^{\lambda-1} e^{-\left(\frac{y-\mu}{x}\right)^{\lambda}} dy$$

$$= \mu \frac{\lambda}{x^{\lambda}} \int_{0}^{\infty} \left(xz^{\frac{1}{\lambda}}\right)^{\lambda-1} e^{-z} \frac{x}{\lambda} z^{\frac{1}{\lambda}-1} dz$$

$$= \mu \int_{0}^{\infty} e^{-z}\, dz$$

$$= \mu$$

Therefore,

$$E(Y/X = x) = x\Gamma_{\left(1+\frac{1}{\lambda}\right)} + \mu \qquad (2)$$

Equation (2) is a regression model which explains the change in the response variable Y for a given value of the predictor variable $X = x$. Although the model is non-linear in the parameter λ, it can be linearized by a change of parameter, say, $\delta = \Gamma_{\left(1+\frac{1}{\lambda}\right)}$; and, replacing $E(Y/X = x)$ by y gives the model

$$y_i = \mu + \delta x_i + \epsilon_i \tag{3}$$

Equation (3) assumes deterministic relationship between the predictor variable x_i and the response variable y_i, perturbed by the disturbance term. However, this relationship results from the addition, into equation (2), of the disturbances, ϵ_i, which summarizes the effect on the response variable of some other factors that have not been included explicitly in the analysis (Pollock, 2014). The assumption is made that these additional and unobserved factors are unaffected by the values of the predictor variable, x_i, that are included in the systematic part of the model. Thus, the $\epsilon_i's$ are regarded as random variables which, under the linear framework, are assumed to be independently and identically distributed across observations. In the present analysis, this strong assumption is relaxed by positing instead that the $\epsilon_i's$ follow a Weibull distribution, reflecting their role as stochastic disturbances in the prediction of $y_i's$.

3. Parameter Estimation

Since the ϵ_i's are assumed to be non-normal, the linear least squares estimation is inappropriate. In addition, since the errors are assumed to be Weibull distributed, modeling their absolute values is suitable. Thus, the least absolute deviation (LAD) technique will be employed in the estimation of parameters of model (3). This technique has some appealing statistical properties, for instance, it reduces the influence of outliers on the estimates of parameters (Dielman, 2009). It is one of the most applied methods in parameter estimation of regression models when the linear assumptions of normality are not tenable; see (Zeckhauser and Thompson, 1970) and (Nevitt and Tam, 1998) for example.

The absolute errors, $u_i = |\epsilon_i|$, from (3) can be expressed as

$$u_i = |y_i - \mu - \delta x_i| \tag{4}$$

The u_i's being absolute deviations are expected to assume a minimum value 0; and since they are Weibull distributed, their probability density function is given by

$$f(u_i) = \frac{\beta}{\alpha^\beta}|y_i - \mu - \delta x_i|^{\beta-1} e^{-\frac{1}{\alpha^\beta}|y_i-\mu-\delta x_i|^\beta} \tag{5}$$

where $|y_i - \mu - \delta x_i| > 0$, $\alpha, \beta > 0$, and $\mu, \delta > 0$

The shape parameter β measures the degree of peak-ness of the density (5) but the scale parameter α has no effect on the maximum likelihood estimates of the parameters of the model (3) as far as the family of densities of the type (5) is concern; μ is the location parameter – which measures the minimum value of

y_i, and δ is the slope of the regression equation (3). Thus, the model (5) can be rewritten as a standard Weibull distribution ($\alpha = 1$) of the absolute errors (4) given by

$$f(u_i) = \beta|y_i - \mu - \delta x_i|^{\beta-1}e^{-|y_i-\mu-\delta x_i|^{\beta}} \tag{6}$$

Given the observations (x_i, y_i), the log-likelihood function of the parameters of the density function (6) is obtained as

$$L\left({}^{\mu,\delta,\beta}/_{x_i,y_i}\right) = n\ln\beta + (\beta-1)\sum_{i=1}^{n}\ln|y_i - \mu - \delta x_i| - \sum_{i=1}^{n}|y_i - \mu - \delta x_i|^{\beta} \tag{7}$$

The first partial derivatives of the log-likelihood function (7), simply $\frac{\partial L}{\partial .}$, with respect to the parameters, gives the system of equations

$$\begin{aligned} -(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-1}k_i + \beta\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}k_i &= 0 \\ -(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-1}x_ik_i + \beta\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}x_ik_i &= 0 \\ \frac{n}{\beta} + \sum_{i=1}^{n}\ln|y_i-\mu-\delta x_i| - \sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta}\ln|y_i-\mu-\delta x_i| &= 0 \end{aligned} \tag{8}$$

where $k_i = sign(y_i - \mu - \delta x_i)$

Closed-form solution to the system of equations (8) are not obtainable. We resort to the Newton-Raphson iteration procedure given by

$$\hat{B}^{\boldsymbol{j}+1} = \hat{B}^{\boldsymbol{j}} + \rho J^{-1} \times -\boldsymbol{f}, \quad j = 0, 1, 2, 3, \ldots \tag{9}$$

where $B^j = \begin{pmatrix} \mu_j \\ \delta_j \\ \beta_j \end{pmatrix}$, μ_0, $\delta_0$and β_o are the initial values of μ, δ and β in the iteration,

$J = \begin{pmatrix} \frac{\partial f_1}{\partial \mu} & \frac{\partial f_1}{\partial \delta} & \frac{\partial f_1}{\partial \beta} \\ \frac{\partial f_2}{\partial \mu} & \frac{\partial f_2}{\partial \delta} & \frac{\partial f_2}{\partial \beta} \\ \frac{\partial f_3}{\partial \mu} & \frac{\partial f_3}{\partial \delta} & \frac{\partial f_3}{\partial \beta} \end{pmatrix}$, and $\boldsymbol{f} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$ is the vector of the partial derivatives in (8).

The partial derives in the matrix J are derived as

$$\frac{\partial f_1}{\partial \mu} = -(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-2} - \beta(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-2}$$

$$\frac{\partial f_1}{\partial \delta} = -(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-2}\, x_i - \beta(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-2}\, x_i$$

$$\frac{\partial f_1}{\partial \beta} = -\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-1}k_i + \sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}\, k_i + \beta\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}\, k_i\, ln|y_i-\mu-\delta x_i|$$

$$\frac{\partial f_2}{\partial \mu} = \frac{\partial f_1}{\partial \delta}$$

$$\frac{\partial f_2}{\partial \delta} = -(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-2}{x_i}^2 - \beta(\beta-1)\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-2}\, {x_i}^2$$

$$\frac{\partial f_2}{\partial \beta} = -\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{-1}k_i\, x_i + \sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}\, k_i x_i + \beta\sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta-1}\, k_i x_i\, ln|y_i-\mu-\delta x_i|$$

$$\frac{\partial f_3}{\partial \mu} = \frac{\partial f_1}{\partial \beta}, \quad \frac{\partial f_3}{\partial \delta} = \frac{\partial f_2}{\partial \beta}$$

$$\frac{\partial f_3}{\partial \beta} = -\frac{n}{\beta^2} - \sum_{i=1}^{n}|y_i-\mu-\delta x_i|^{\beta}\,(ln|y_i-\mu-\delta x_i|)^2$$

f_1, f_2 and f_3 are the system of equations (8), respectively. Values for $\boldsymbol{f}$and J are obtained using the j^{th} iteration values of the parameters μ, δ and β. The relaxation factor ρ, typically set to 0.5 stabilizes convergence in the iterative procedure (9).

The iterative procedure (9) is implemented in R statistical programming language using the program, WeiReg.

4. Test for Best Fit of the Regression Model

Unlike the linear regression model which assumes ϵ_i is normally distributed with mean zero and variance σ^2, the regression model in this study assumes that $|\epsilon_i| \sim Weibull(1, \beta)$, thus the name Weibull regression model. To test for the fitness of the new model to data, we use the likelihood ratio test to compare the log-likelihood function value of the new model to that of the linear regression model. The appropriate hypothesis for test is given by

$$H_0: \epsilon_i \sim N(0, \sigma^2) \ vs \ H_1: |\epsilon_i| \sim Weibull(1, \beta)$$

The test statistic is given by

$$LR = -2(Ln - Lw) \tag{10}$$

where Ln is the log-likelihood function value of the linear regression model and Lw is the log-likelihood function value of the Weibull regression model; LR is chi-squared distributed with 1 degree of freedom.

5. Model Selection

The model selection criteria provide a generalized framework for evaluating efficiency and goodness of fit. One of the most widely used approaches is the Akaike Information Criterion (AIC), which balances model fit against model complexity. It is defined as

$$AIC = -2L(\hat{\theta}) + 2k, \tag{11}$$

where $L(\hat{\theta})$ is the maximized likelihood function for the model under consideration, and k is the number of estimated parameters. Lower AIC value indicates a better trade-off between fit and parsimony.

In practice, the difference in AIC values, $\triangle AIC$, between two competing models is often interpreted as the relative evidence of a better fit in favor of one model over the other (Burnham and Anderson, 2002).

6. Application

6.1 The Data

The data used for illustration are the wind speed datasets recorded at 9 a. m. (X) and 3 p. m. (Y) at Christmas Island, Australia. The data were collected as daily records from October 1, 2018 to March 25, 2020. To clean the data, days with missing records for both or one of the time points were excluded. Thus, 517 of the 542 daily records had the full bivariate datasets. Also, the original data in km/h were converted to m/s by multiplying each entry by a factor of 0.278. The data link is www.bom.gov.au>climate>datasets.

6.2 Parameter Estimates

Parameter estimates using the iterative equation (9) in R produce the following results along with the linear regression output.

Table 1 Parameter estimates of Weibull and linear regression models of Wind speeds at 9 a. m. and 3 p. m. at Christmas Island, Australia

	Weibull Regression			**Linear Regression**	
	μ	δ	β	a	b
Estimate	1.2037	0.7203	1.3046	1.2430	0.7170
Standard error	0.0159	0.0025	0.0476	0.1229	0.0231
t statistic	75.535	283.72	27.427	10.114	31.039
$\boldsymbol{p-value}$	0.0000	0.0000	0.0000	0.0000	0.0000

It is obvious from the parameter estimates in Table 1 that both regression methods have fitted the wind speed datasets properly, considering the minimum values of the standard errors of the parameter estimates. In addition, the parameter estimates are shown to be all significant considering their p-values.

6.3 Test of best fit

The test for best fit results produce the value of $LR = 674.82$ and $p - value = 0.0000$. This means that the null hypothesis that the data is best described by the linear regression model is rejected, while the alternative hypothesis that the data is best fitted by the Weibull regression model is accepted.

6.4 AIC-Based Model Comparison

The log-likelihood values and corresponding Akaike Information Criterion (AIC) statistics for the fitted regression models are presented in Table 2. These metrics provide an objective basis for comparing the relative adequacy of the Weibull regression model against the linear regression model in capturing the underlying wind speed dynamics.

Table 2 Comparison of Weibull regression and linear regression models

	Weibull regression	**Linear regression**
Log-likelihood	- 357.49	- 694.90
AIC	721.0	1393.8

From the results, the Weibull regression model yields both a substantially higher log-likelihood (–357.49 compared to –694.90) and a lower AIC value (721.0 compared to 1393.8). Since AIC incorporates a penalty for model complexity while rewarding better fit, the marked reduction in AIC indicates a clear improvement in explanatory power without undue overfitting.

This evidence strongly supports the superiority of the Weibull regression approach over the linear regression model, which is constrained by the assumption of normally distributed errors. In particular, the lower AIC value signifies that the Weibull regression more efficiently captures the probabilistic structure of the wind speed data, aligning with theoretical expectations that wind speed distributions often exhibit skewness and heavy tails that normal-based models fail to accommodate (Coles, 2001; Burnham & Anderson, 2002).

7. Discussion

The results of this study demonstrate that the proposed regression model with a Weibull error term provides a robust alternative to the linear linear regression framework. By adopting the least absolute deviation (LAD) estimation technique, the model reduces the influence of outliers on parameter estimates and offers improved efficiency in contexts where normality assumptions are untenable (Zeckhauser & Thompson, 1970; Dielman, 2009). This feature is particularly valuable for environmental datasets such as wind speed, which often display skewness and heavy tails that normal-based models cannot adequately capture (Coles, 2001; Baissa & Rainey, 2020).

The superiority of the Weibull regression model over the linear regression model is supported by multiple lines of evidence. First, the likelihood ratio test strongly rejects the null hypothesis of normal errors in favor of Weibull-distributed disturbances. Second, the AIC values provide compelling support for the Weibull regression, with the markedly lower AIC (721.0 versus 1393.8) indicating a substantially better balance between fit and parsimony (Burnham & Anderson, 2002). This aligns with findings in the broader literature where flexible error structures improve inference and predictive accuracy (Huber, 1981; Lindsey, 1997; Manning *et al*., 2003).

Taken together, these findings suggest that the Weibull regression model is not only statistically efficient but also practically relevant for modeling skewed, non-negative outcomes in environmental and life sciences applications. By extending the range of error distributions considered in regression analysis, this approach contributes to the growing literature on robust methods that accommodate departures from linear normality assumptions (Tiku *et al*., 1986; Ganguly, 2014).

8 Conclusion

This study has shown that the Weibull regression model, estimated via least absolute deviations, provides a more robust and efficient framework than the linear regression model for analyzing wind speed data. The substantially lower AIC and higher log-likelihood values highlight its superior fit, while accommodating the skewed and heavy-tailed nature of the data. These findings underscore the importance of adopting flexible error distributions in regression analysis, particularly in environmental applications where normality assumptions are often violated.

References

Aldrich, J. (2005). Fisher and Regression, *Statistical Science,* vol. 20, No 4, 401 – 417, https://DOI:10.1214/088342305000000331 © Institute of Mathematical Statistics

Baissa, D. K. and Rainey, C. When BLUE is not best: Non-normal errors and the linear models. *Political Science Research and Methods*, (2020), 8, 136 –148. https://doi:10.1017/psrm.2018.34

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer.

Dielman, T. E. (2009). Least Absolute Value verse Least Squares Estimation and Inference Procedures in Regression Models with Asymmetric Error Distributions, *Journal of Modern Applied Statistical Methods*, vol. 8, Issue 1, Article 13, pp 147 – 160 @ http://digitalcommons.wayne.edu/jmasm/vol8iss1/13

Ganguly, S. S. (2014). Robust Regression Analysis for Non-Normal Situations Under Symmetric Distributions Arising in Medical Research, *Journal of Modern Applied Statistical Methods*, vol. 13, Issue 1, Article 29, @ http://digitalcommons.wayne.edu/jmasm/vol13iss1/29

Gongsin I. E. and Saporu F. W. O. (2020) A bivariate conditional Weibull distribution with application, *Afrika Matematika*, 31:565–583. https://doi.org/10.1007/s13370-019-00742-8

Gujarati, D. N., Porter, D. C. and Gunasekar, S. (2012). *Basic Econometrics*, Fifth Edition, Tata McGraw Hill Education Private Limited, New Delhi.

Huber, P. J. (1981). *Robust Statistics,* John Wiley, New York.

Lindsey, K. J. (1997). *Applying Generalized Linear Models*, © 1997 Springer-Verlag, New York, Inc.

Manning, W. G., Basu, A, and Mullahy, J. (2003). Generalized Modelling Approaches to Risk Assessments of Skewed Outcomes Data. *Technical Working paper 293*, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, M A 02138, October 2003, @ http://www.nber.org/papers/T0293

Nevitt, J. and Tam, H. P. (1998). A Comparison of Robust and Nonparametric Estimators Under the Simple Linear Regression Model, *Multiple Linear Regression Viewpoints*, Vol. 25, pp 54 – 69.

Pollock, D. S. G. 2014, *Econometrics: A Historical Guide for the Uninitiated,* Working Paper No. 14/05, February 2014, University of Leicester, UK.

The MIT Press, http://www.jstor.org/stable/1926296

Tiku, M. L., Tau, W. Y. and Balakrishnan N. (1986). *Robust Inference*, Marcel Dekker, New York.

Wooldridge, J. M. 2013, *Introductory Econometrics: A Modern Approach*. 5th ed. Mason, Ohio: South-Western Cengage Learning.

Zeckhauser, R. and Thompson, M. (1970). Linear Regression with Non-Normal Error Terms, *The Review of Economics and Statistics*, Vol. 52, No. 3 (Aug., 1970), pp. 280-286