Global Scientific JOURNALS
GSJ: Volume 12, Issue 5, May 2024, Online: ISSN 2320-9186
www.globalscientificjournal.com

Residual Analysis For Kumaraswamy Weibull Regression Model with Application.

Ahmed Abdalla , Salah M. Mohamed[2], Amal Mohamed
[2] *Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research (FSSR), Cairo University, Egypt,*
***Email Address**:[1] corresponding Author, ahamed12763@gmail.com..*

Abstract
In this research paper, we propose a new regression for the Kumar-Samy-Weibull distribution, The coefficients of the proposed regression model were estimated using the Maxim method ,in the finally the proposed model was applied to a set of real data. The proposed model was compared to some models through some statistical criteria

Keywords: *Log Kumaraswamy Weibull***,** *Pearson Residuals***,**Deviance component Residual, *The martingale residual,*

1. INTRODUCTION AND PRELIMINARIES

Regression plays an important role in providing solutions to a number of problems, probability distributions, that is, the tools that represent data, so we note that many studies have presented regression for a number of distributions. **[8],** introduced log beta log-logistic regression model based on the beta log-logistic distribution][1**]** Introduced the gamma regression model**[8])** introduced Kumaraswamy Lindley regression model **[9]** The xgamma Family: Censored Regression Modeling[2] introduced The beta regression with application **[3]** introduced a novel approach for the estimation of quantiles regression

The article is organized as follows: In Section (2), Definition Kumaraswamy family Distribution, Definition of Weibull distribution, The quantile function (qf), The Survival function and Hazard Rate Function. H (x) In Section(3) we suggested a log Kumaraswamy weibull regression model of location-scale form, Using the maximum likelihood method to estimate the parameters and construct the observed information matrix. In Section (4) ,type of residual analysis are presented in section (5)

Definition 1.1. ***Kumaraswamy family Distribution***:

Generalized distributions are very important in the field of probability distributions, as these generalized distributions contain many mathematical properties that make the distribution more flexible. Generalized distributions depend on two things, namely the probability density function of the distribution (pdf) and the cumulative distribution function (CDF) defined the K-G family by the pdf and CDF given by

,

$$f(x) = abg(x)G(x)^{a-1}\left[1-G(x)^{a}\right]^{b-1} \quad (1.1)$$

where g(x) is density function for distribution and G(x) corresponding cumulative functic he base line.

,

$$F(x) = 1-\left[1-G(x)^{a}\right]^{b} \quad (1.2)$$

Respectively, where a and b are two additional positive shape parameters. Clearly, for a = b = 1, we obtain the baseline distribution. The additional parameters a and b aim to govern skewness and tail weight of the generated distribution.

2.2-Definition of Weibull distribution:

The Weibull distribution is play very important role for modeling lifetime data in many filed for example medicine, biology, engineering and finance, a very small amount of the massive applications of Weibull model including, The density function of the two parameters Weibull model is given by

,

$$f(x) = \frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} \exp(-(\frac{x}{\beta})^{\alpha}) \quad (1.3)$$

The cumulative distribution function corresponding with the distribution

,

$$F(x) = 1-\exp(-(\frac{x}{\beta})^{\alpha}) \quad (1.5)$$

2.3- Definition of (GKW) Distribution:

,

$$f(x) = ab\,\frac{\alpha}{\beta^{\alpha}} x^{\alpha-1} \exp(-(\frac{x}{\beta})^{\alpha})\left[1-\exp(-(\frac{x}{\beta})^{\alpha})\right]^{a-1}\left[1-\left[1-\exp(-(\frac{x}{\beta})^{\alpha})\right]^{a}\right]^{b-1} \quad (1.6)$$

The CDF of the (GKW) distribution can be written as

$$F(x) = 1-\left[1-\left[1-\exp(-(\frac{x}{\beta})^{\alpha})\right]^{a}\right]^{b} \quad (1.7)$$

2.4-The quantile function (qf):

The quantile function (qf) is considered one of the most important in the probability distributions, as it is possible to obtain from it the cumulative distribution function. It is also possible through this function to study some mathematical properties of the distribution. After this importance, The quantile function (qf) must be calculated for the GK-Weibull

$$Q(a,b,\alpha,\beta)=\beta\left[-lin\left(1-\left\{1-(1-u)^{\frac{1}{b}}\right\}^{\frac{1}{a}}\right)\right]^{b} \quad (1.8)$$

2.5- The Survival function:

The survival function of X is defined as Substituting in the cumulative distribution function, we get

$$s(x)=\left[1-\left[1-\exp(-(\frac{x}{\beta})^{\alpha})\right]^{a}\right]^{b} \quad (1.9)$$

3-Log G-kw Regression Model:

Explanatory factors must be included in models when they have an impact on the response variable X=f(y). Parametric models are frequently used to include the information about the explanatory variables. Regression models that fit lifetime data well typically produce more accurate estimates of the important quantities. When explanatory factors are truly necessary but are not included in the model, a sizable portion of the response variable's variability will be left behind as residual. Many regression models have recently been published in the literature that take location into account. Cordeiro et al. (2016), for instance, introduced a new class of survival regression models that take the log-gamma extended into consideration. The beta generalised half-normal geometric regression model was introduced by Ramires et al.(2013). Suppose x is a random variable following the G-kw density function (3.4) and y is defined $y = logx$,$\beta = e^{\mu}$,$\alpha = \frac{1}{\sigma}$, It is easy to verify that the density function of y reduces to

$$f(y)=J.f^{-1}(x) \quad (1.10)$$

To obtain the value of the Jacobean transformation, the differential hypothesis with respect to x we get

$$J=(dx/dy)=x \quad (1.11)$$

By substituting in equation (1.6) & (1.7) for the value of the probability density function and the value of the Jacobi, we get the following

,

$$f(x)=\frac{ab}{\sigma}\exp(\frac{y-\mu}{\sigma}).\exp(-\exp(\frac{y-\mu}{\sigma}))\left[1-\exp(-\exp(\frac{y-\mu}{\sigma}))\right]^{a-1}\left[1-\left[1-\exp(-\exp(\frac{y-\mu}{\sigma})\right]^{a}\right]^{b-1}$$

where $-\infty < \mu < \infty$, $\sigma > 0$. We refer to equation (3.15) as the GKW distribution, say $f(\boldsymbol{y}) \sim$ GKW $(\sigma, \mu, \mathbf{a}, b)$, where $\mu \in$R is the location parameter, σ >0 is the scale parameter, corresponding CDF and survival function respectively are :

$$F(x)=1-\left[1-\left[1-\exp(-\exp(\frac{y-\mu}{\sigma}))\right]^{a}\right]^{b} \quad (1.12)$$

4-Maximum Likelihood Estimation:

Consider a collection of **n** independent observations $(x_1, y_1),..., (x_n, y_n)$, where each random re sponse is specified by $y_i = \min(\log(x_i), \log(c_i))$. Since the observed lives and censoring timings are independent, we assume non-informative censoring times are independent, Let F and C be the sets of individuals for which yi is the log-lifetime, respectively. Conventional likelihood estimation techniques can be applied here. The log-likelihood function for the vector of parameters The total log-likelihood function for $L(\theta)$ reduces to

$$\ln L(\theta) = \sum_{i=1}^{n} \delta_i \log f(y) + \sum_{i=1}^{n} (1-\delta_i) \log s(y) \quad (1.13)$$

Substituting in equation (1.13) for the value of the density function and the cumulative distribution function, we get the following: $\ln L(\theta) = k_5 + K_6$

Where $K_7 = \sum_{i\in F}^{n} Log(f(y_i))$ $K_8 = \sum_{i\in c}^{n} log(S(y_i))$ To calculate a value of K_7 in order to simplify the density function

$$k_7 = \sum_{i=1}^{n} \delta_i \log\left(\frac{ab}{\sigma} \exp(\frac{y-\mu}{\sigma}).\exp(-\exp(\frac{y-\mu}{\sigma}))\left[1-\exp(-\exp(\frac{y-\mu}{\sigma}))\right]^{a-1}\left[1-\left[1-\exp(-\exp(\frac{y-\mu}{\sigma})\right]^{a}\right]^{b-1}\right)$$

$$k_8 = \sum_{i=1}^{n} (1-\delta_i) \log\left[\left[1-\left[1-\exp(-\exp(\frac{y-\mu}{\sigma}))\right]^{a}\right]^{b}\right]$$

Estimation of regression coefficients. Differentiation of the total log-likelihood function with respect to regression coefficients

$$\frac{\partial \ln L(\theta)}{\partial \beta_0} = \frac{\partial k_7}{\partial \beta_0} + \frac{\partial K_8}{\partial \beta_0} \quad (1.14)$$

$$\frac{\partial \ln L(\theta)}{\partial \beta_1} = \frac{\partial k_7}{\partial \beta_1} + \frac{\partial K_8}{\partial \beta_1} \quad (1.15)$$

5-Residual analysis:

We need a tool to check the assumptions after the model is constructed so we defined for types of residuals: the martingale residual, the deviance component residual, cox residual and Pearson Residual, Cox and snell Residual, now I well explain all concepts.(Fernandez, G. C. 1992)

5.1-The martingale residual, is much used in the counting process These residuals are asymmetric and take maximum values (+1) and minimum values (-∞) we defined the martingale residual as

$$r_{M_i} = \delta_i + \log(S(y))$$

$$r_{M_i} = \delta_i + \log(\left[\left[1-\left[1-\exp(-\exp(\frac{y-\beta_0 - x\,\beta_1}{\sigma}))\right]^a\right]^b\right])\delta_i = 0,1 \qquad (1.15)$$

5.2-Deviance component Residual:

This residue was suggested to make the martingale residual more symmetric around zero. The deviance component for the parametric regression model is given

$$r_{M_i} = sign(r_{M_i}).\delta_i \left[-2\left[r_{M_i} + \log(\left[1-r_{M_i}\right]\right]\right]^{\frac{1}{2}}) \qquad (1.16)$$

Where r_{Mi} is the martingale residual. * sign () function is a function that drives the (***+1***) values if the argument is positive and (***-1***) is negative The deviance component residual for the Weibull model is given by

5.3-Pearson Residuals: are used to detect outliers It depends on the idea of subtracting the mean and dividing by the standard deviation. The Pearson Residual knows the bounded regression of the inverse of the exponential distribution as follows:

$$r = \frac{y-\mu}{\sqrt{\text{var}(y)}} \qquad (1.17)$$

5.4-Cox and snell Residual:

Cox and snell (1968) residual defined as follows:

$$e_i = -\ln(1-(1-e^{-e^{\frac{y-\mu}{\sigma}}})^a)^b \qquad (1.18)$$

Substituting in equation No.(27) for the value of the Cumulative Function, we get the following:

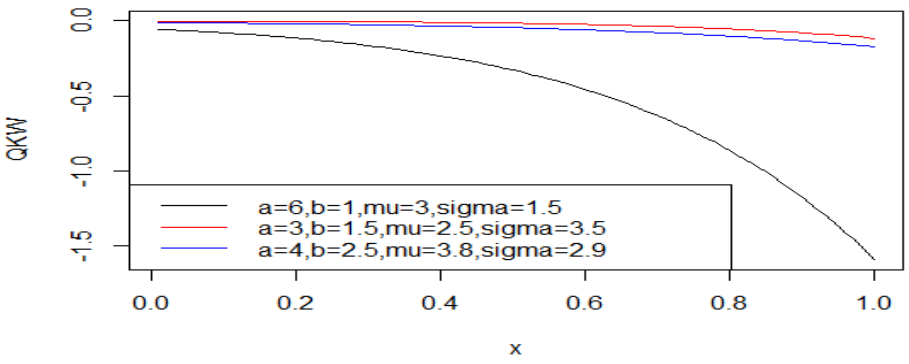

Figure (1)The **Cox and snell Residual** for kumaraswamy weibull regression model at different value

5.5-Global influence:

Does the diagnostic impact depend on case reduction, one of the tools introduced by Cook to perform sensitivity analysis? It depends on the effect on parameter estimates of removing observations. This method compares $\theta_i$and $\vartheta_i - 1$. where γ_ii is the MLE when ith observations are removed from the original data. Then the ith case can be considered as influential observation if $\vartheta_i - 1$ is far away from θ_i This methodology was conducted in many statistical models, The case deletion model for the GK-Weibull

2. MAIN RESULTS

6-Sumlation Study:

In this section, a simulation study is given to evaluate the performance of coefficients for the proposed regression model, According to and following steps using Monte Carlo Simulation.

1- Taking initial value (a=2,b=2, β_0=0.6, β_1=0.6, σ=1)

2- Generate $y \sim W(\mu, \sigma)$ **where** $\mu = \beta_0 + x\beta_1$

3- Generate $x \sim U(0,1)$

4- Generate $z \sim U(0,1)$

5- Taking samples (20, 50, 100)

6-The simulation replication is N = 1 000.

7-For each generated sample sizes, the biases ,(AS)and (MSE) are evaluates at level The three censoring rates (20%, 30%, 40%)

The simulation results are reported in Table (1),Table(2) and Table (3) . As seen from the results, the estimated biases, average of estimates (AS) and mean square error (MSEs) are near the desired value, zero

Table 2.1. Simulation Study Results of (K-W) Regression model

	N=20			N=50			N=100		
	AS	bias	MSE	AS	bias	MSE	AS	bias	MSE
a	2.423	0.423	0.726	2.312	0.312	0.623	2.021	0.021	0.041
b	2.324	0.324	0.812	2.223	0.233	0.423	2.001	0.001	0.021
β_0	0.812	0.212	0.421	0.742	0.224	0.321	0.611	0.012	0.002
β_1	0.923	0.323	0.942	0.723	0.123	0.523	0.512	0.112	0.121
σ	1.231	0.231	0.421	1.121	0.121	0.144	1.002	0.002	0004

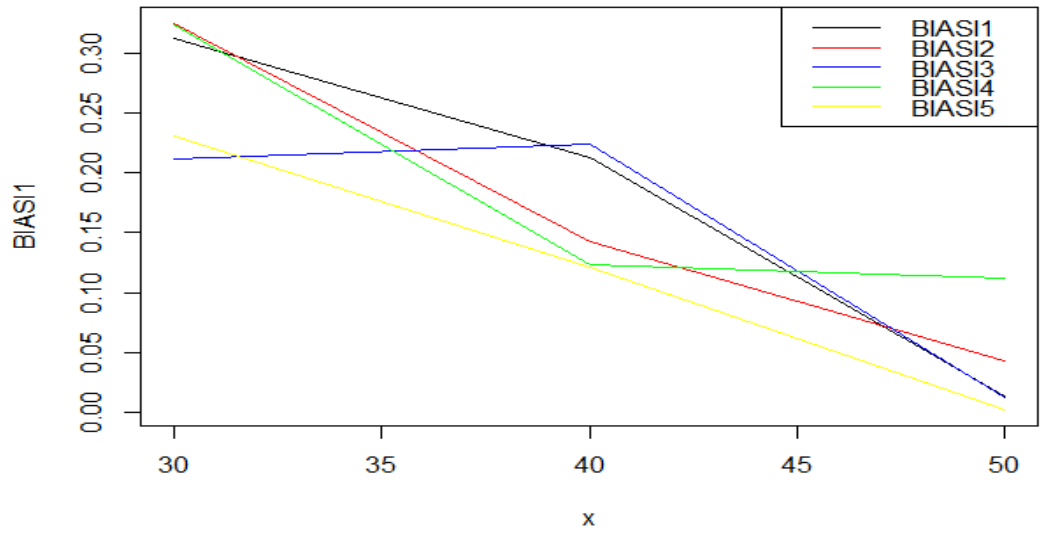

Figure (2) The bias kumaraswamy weibull regression model at different value

Table 2.2. Simulation Study Results of (K-W) Regression model

Censoring rate 0.30	N=20			N=50			N=100		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	2.312	0.312	0.924	2.213	0.213	0.421	2.013	0.013	0.010
b	2.325	0.325	0.923	2.142	0.142	0.196	2.043	0.043	0.016
β_0	0.742	0.142	0.163	0.712	0.012	0.162	0.613	0.010	0.014

β_1	0.843	0.243	0.423	0.753	0.152	0.225	0.643	0.043	0.013
σ	1.240	0.240	0.311	1.231	0.231	0.214	1.025	0.025	0.062

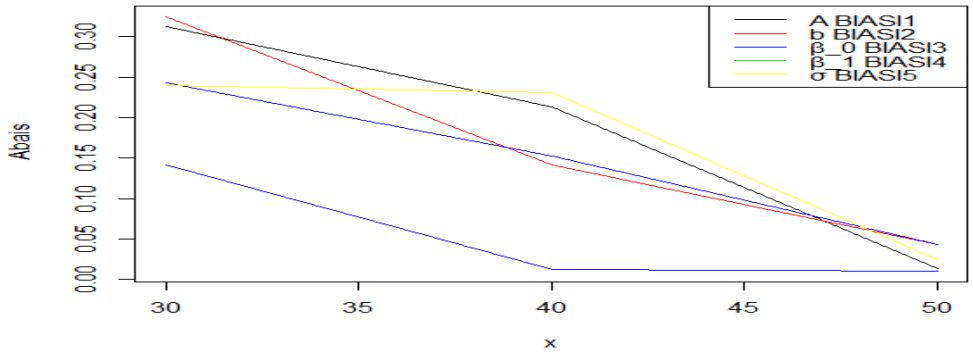

Figure (3) The bias kumaraswamy weibull regression model at different value rate 0.3

Table 2.3 Simulation Study Results of (K-W) Regression model

Censoring rate 0.40	N=20			N=50			N=100		
	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
a	2.456	0.456	0.562	2.342	0.342	0.413	2.023	0.023	0.004
b	2.732	0.723	0.492	2.534	0.543	0.253	2.043	0.043	0.061
β_0	0.842	0.242	0.423	0.624	0.024	0.412	0.321	0.011	0.023
β_1	0.795	0.195	0.423	0.675	0.075	0.049	0.657	0.057	0.034
σ	1.423	0.423	0.162	1.125	0.125	0.025	1.025	0.025	0.004

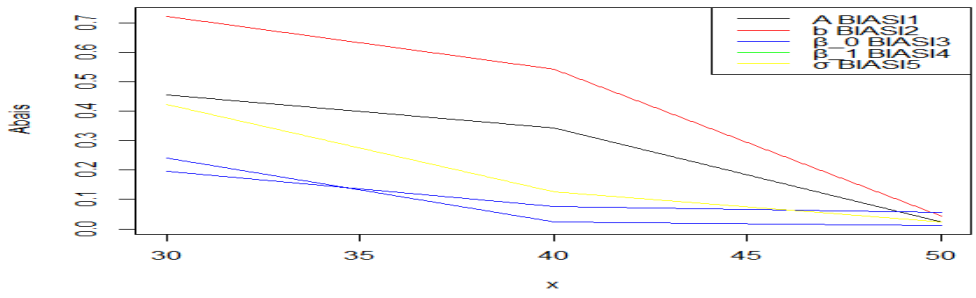

Figure (4)The bias kumaraswamy weibull regression model at different value rate 0.40

-Real data

The real data in this research consisted of a relationship between two variables, one of which is an independent variable and the other is a dependent variable, where the dependent variable represents the risk coverage rate(**y2**) and the independent variable represents the return on investment.**(x23)**

Table 2.4 descriptive statistics for the independent variable and the dependent variable

Variable	*Min.*	*Mean*	*Max*	*1st Qu*	*3rd Qu*
the risk coverage rate(y2)	10.18	339.77	1196.43	60.53	615.33
the return on investment.(x23)	-0.5878	0.21008	0.77650	-0.050	0.52390

By extrapolating the previous table, we notice that the maximum value of the independent variable was (0.77650) while the minimum value of the independent variable was (-0.5878), and the arithmetic mean of the independent variable was (0.21008), While the maximum value of the dependent variable was (1196.43) and the minimum value of the dependent variable was (10.18) and the arithmetic mean of the same variable was (339.77)

Goodness- of – Fit Test of Real Data

Table (2.5) Goodness- of – Fit Test of Real D

Goodness of fit statistics	**kumrsawamy weibull**	**weibull**	**MG-weibull**
Kolmogorov- sminrov	**0.07467584**	0.0989061	0.094695
Cramer- von statistic	**0.01312341**	0.0234912	0.097667
Anderson- darling	**0.13707113**	0.1734884	0.624825

From the previous table, we notice the following that the value of some statistical criteria, such as the Kolmogorov- sminrov test, the Cramer- von statistic test, and Anderson- darling, for the kumaswamyweibull distribution is samller than the rest of the other distributions, and thus the data represents the kumrsawamy Weibull distribution.

4.5 Fit of The Regression Model for kumaraswamy weibull

In this section, the regression parameters of the proposed model are estimated and the proposed model is compared to some other models through some statistical criteria.

Table (2.6): Estimation of Parameters for the kumaraswamy weibull Regression

Regression model	$\widehat{\boldsymbol{\beta_0}}$	$\widehat{\boldsymbol{\beta_1}}$	*HQIC*	*BIC*	*AIC*	*R^2*
Kumaraswamy weibull	0.42352	0.6234	132.2815	125.3456	124.5423	0.8986
kumaraswamy Lindley	0.03465	0.5234	134.8452	130.7256	128.4213	0.4239
MG weibull	0.34214	0.4652	138.232	139.341	136.423	0.3492

we notice that each of the statistical criteria for (AIC,BIC,HQIC) the Kumarsumy-weibull regression is smaller than the rest of the same criteria for the other regressions. Thus, the Kumarasumy weibulle regression is the best regression. The same is true for the coefficient of determination. We note that the coefficient of determination for the Kumarasumy-webull regression is larger than the rest of the other regressions. Thus, the Kumarasumy weibulle regression is the best regression

Conclusions

In this research paper, we suggested the Kumaraswamy Weibull regression model which is convenient for lifetime data. Using maximum likelihood estimation, to estimate the parameters. and compared the suggested regression model with sub-models; the Lindley regression model, Weibull regression model, and gamma regression model by using the BIC, and AIC criteria. Moreover, the regression models were applied to real data, A survival time prediction, We observed that the proposed regression model outperforms sub-mode

References

1- Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., & Taha, I. M. (2022**).** A new two-parameter estimator for beta regression model: method, simulation, and application. Frontiers in Applied Mathematics and Statistics, Vol7, No3, 242-252..

2- Amin, M., Akram, M. N., & Amanullah, M. (2022). On the James-Stein estimator for the Poisson regression model. *Communications in Statistics-Simulation and Computation*, vol 51, No 10, 5596-5608..

3-Cordeiro, G. M., Altun, E., Korkmaz, M. C., Pescim, R. R., Afify, A. Z., & Yousof, H. M. (2020). The xgamma family: Censored regression modelling and applications. *Revstat-Statistical Journal*, VOL18,N5, 593-612.

4- Cordeiro, G. M., & Simas, A. B. (2009**).** The distribution of Pearson residuals in generalized linear models. Computational statistics & data analysis, vol*53*,N9, 3397-3411.

5- Ortega, E. M., Barriga, G. D., Hashimoto, E. M., Cancho, V. G., and Cordeiro, G. M. (2016). A New Class of Survival Regression Models with Cure Fraction. *Journal of Data Science*, *12*.

6- Loynes, R. M. (1969). On Cox and Snell's general definition of residuals. *Journal* of the Royal Statistical Society Series B: Statistical Methodology, Vol31,N2, pp103-106

7-Mahmoud, M. R., El-Sheikh, A. A., Morad, N. A., & Ahmad, M. A. (2015). Log-beta log-logistic regression model. Int. J. Sci. Basic Appl., Res.(IJSBAR*)*,

8-Abdelmoezz, S., & Mohamed, S. M. (2021). The Kumaraswamy Lindley Regression Model with Application on the Egyptian Stock Exchange: Numerical study, Regression model. *Jurnal Matematika, Statistika dan Komputasi*, vol18,N1, 1-11.

9-Ramires, T., Ortega, E., Cordeiro, G., & Hamedani, G. (2013). The beta generalized half-normal geometric distribution. Studia Scientiarum Mathematicarum Hungarica, Vol50,N4, 523-554.)

10- Fernandez, G. C. (1992). Residual analysis and data transformations: important tools in statistical analysis. HortScience, vo27.N(4), pp297-300.