DOI: http://doi.org/10.4038/sljastats.v23i1.8066

Sri Lankan Journal of Applied Statistics, Vol (23-1)

Unit Gamma/Gompertz Quantile Regression with Applications to Skewed Data

M.H.B. Mustapha* **and S. Nasiru**

Department of Statistics, School of Mathematical Sciences,
C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana

*Corresponding author: mhmustapha.stu@cktutas.edu.gh

Received: 16^{th} March 2022/ Revised: 10^{th} August 2022/ Published: 31^{st} August 2022

ABSTRACT

In this study, new unit quantile regression model, called the Unit Gamma/ Gompertz quantile regression for bounded responses is developed by re-parameterizing the Unit Gamma/Gompertz distribution. To estimate the parameters of the new quantile regression model, the maximum likelihood approach is used to develop estimators for the parameters. Monte Carlo simulations are used to test the consistency of the maximum likelihood estimators for the parameters of the new quantile regression model. The application of the new quantile regression model is illustrated using three real life datasets and the results revealed that the Unit Gamma/Gompertz quantile regression performs better than the beta regression model when the unit response variable has skewed observations and outliers.

Keywords: Distribution, Probability density function, re-parameterizing, Maximum likelihood estimation, Cox-Snell residuals.

1 Introduction

In recent years, the development of unit distribution has accelerated significantly. The emphasis is on modeling a wide range of events using data with values within the unit interval, such as proportions, probabilities, and percentages, among others. The development of parametric, semi-parametric, and non-parametric regression models is also in high demand in applied fields for the analysis of such data, and it is growing year after year. The conditional mean function is used in classical linear regression analysis to estimate the mean of the response for each fixed value of the explanatory variables. The mean is one of the most important measures of central tendency, but it tells us

very little about the tail area of the distribution. If researchers' primary goal is to examine the entire conditional distribution, then the classical regression model cannot be extended to non-central locations.
The beta regression model is the most commonly used model where variables are linked with responses that are continuously measured on the unit interval due to its flexibility and clear parameter interpretation. The regression parameters in this model are interpretable in terms of the mean, as the model is naturally heteroscedastic and can accommodate asymmetries. Although the beta distribution is adaptable to fit data on the unit interval, additional unit interval regression models have been presented in the literature. The most popular are the simplex regression model (Barndorff-Nielson and Jørgensen, 1991), Kumaraswamy regression model (Mitnik et al., 2013), log-Lindley regression model (Gómez-Déniz et al., 2016), Johnson SB regression model (Lemonte et al., 2016), Unit Gamma (UG) regression model (Mousa et al., 2016), unit-Logistic regression model (Paz, 2017), unit-Lindley regression model (Mzucheli et al., 2019), Flexible quasi-beta regression (Bonat et al., 2019), unit-improved second-degree Lindley regression (Altun and Cordeiro, 2020), unit generalized half normal regression model (Korkmaz, 2020) and log-weighted exponential regression model (Altun, 2021) among others. In contrast to the classical regression model, the quantile regression model is robust because the mean is sensitive to outliers and is delicately impacted by skewed data. Thus, quantile regression modeling is superior to classical mean regression modeling if the response variable contains outliers (Mazucheli et al., 2021). Some quantile regression models for bounded response that are in literature are: the Bayesian quantile regression model (Santos et al., 2015), new Kumaraswamy quantile parametric mixed regression model (Bayes et al., 2017), unit-Weibull quantile regression (Mazucheli et al., 2020), unit Birnbaum–Saunders quantile regression model (Mazucheli et al., 2021), transmuted unit Rayleigh (TUR) quantile regression model (Korkmaz et al., 2021). unit Burr-XII quantile regression model (Korkmaz et al., 2021), exponential geometric quantile regression model (Jodra et al., 2021) and log exponential-power quantile regression model (Korkmaz et al., 2021).
In this study, we formulated a quantile regression model considering a re-parameterization of the unit-Gamma/Gompertz (UG/G) distribution in terms of its quantile function. By re-parameterizing the UG/G distribution in terms of ρ^{th} quantile, one gets the interpretation of its location parameter as being the ρ^{th} quantile of the distribution. The idea of re-parameterizing was inspired by Mazucheli et al. (2020), Mitnik et al. (2013) and Bayes et al. (2012).

The rest of this study is structured as follows. Section 2 describes the UG/G distribution, the re-parameterized UG/G Distribution and some of its key features. Section 3 describes the UG/G quantile regression model, parameter estimation and the Cox-Snell residual diagnostic. Section 4 conducts a simulation analysis to examine the maximum likelihood estimators' finite sample behavior. Section 5 offers three real-life applications that make use of the proposed quantile regression model and the beta regression model, as well as a residual analysis to evaluate the efficacy of the new model. The study concludes with some discussions, conclusions and recommendations for future study in section 6.

2 Unit Gamma/Gompertz Distribution and Re-parameterized Unit Gamma/Gompertz Distribution

2.1 Unit Gamma/Gompertz Distribution

Bantan et al. (2021) presented the UG/G distribution by inverse-exponentially converting the Gamma/Gompertz distribution established by Bemmaor and Glady (2012). The probability distribution function (PDF) and the cumulative distribution function (CDF) of the UG/G distribution are given by

$$f_{UG/G}(y;\alpha,\vartheta,\theta) = \frac{\alpha\vartheta\theta^{\vartheta}y^{\alpha\vartheta-1}}{(1+(\theta-1)y^{\alpha})^{\vartheta+1}}, y \in (0,1), \tag{1}$$

and

$$F_{UG/G}(y;\alpha,\vartheta,\theta) = \frac{\theta^{\vartheta}}{(\theta-1+y^{-\alpha})^{\vartheta}}, y \in (0,1), \tag{2}$$

respectively, where $\alpha > 0$, $\vartheta > 0$, $\theta > 0$ are shape parameters.
The corresponding hazard rate function of the UG/G distribution is given by

$$h_{UG/G}(y;\alpha,\vartheta,\theta) = \frac{\alpha\vartheta\theta^{\vartheta}y^{\alpha\vartheta-1}}{(1+(\theta-1)y^{\alpha})[(1+(\theta-1)y^{\alpha})^{\vartheta}]}, 0 \leq y \leq 1 \tag{3}$$

Also, the quantile function of the UG/G distribution is the inverse of the CDF. It is mathematically expressed as

$$Q_{UG/G}(\rho;\alpha,\vartheta,\theta) = \left(1+\theta\left(\rho^{-\frac{1}{\vartheta}}-1\right)\right)^{-\frac{1}{\alpha}}, \rho \in [0,1]. \tag{4}$$

2.2 Re-parameterized Unit Gamma/Gompertz Distribution

The UG/G distribution is redefined based on its quantile function in this section to enable us study the effect of one or more covariates on a response variable that follows the UG/G distribution. To obtain the re-parameterized density function we make the subject in the quantile function given in equation (4). Let $\tau = Q_{UG/G}(\mu; a, r, \theta)$ in equation (4), then

$$\theta = \frac{\tau^{-\alpha}-1}{\rho^{-\frac{1}{\vartheta}}-1}, \rho \in [0,1]. \tag{5}$$

where ρ is an assume quantile value. Substituting θ in equation (5) into the density function of the UG/G distribution given in equation (1) yields the re-parameterized density function. Hence, the new density function in terms of the quantile is given as

$$f_{UG/G}(y;\alpha,\vartheta,\rho,\tau) = \frac{\alpha\vartheta(\tau^{-\alpha}-1)^{\vartheta}y^{\alpha\vartheta-1}}{(\rho^{-\frac{1}{\vartheta}}-1)^{\vartheta}[1+(\frac{\tau^{-\alpha}-\rho^{-\frac{1}{\vartheta}}}{\rho^{-\frac{1}{\vartheta}}-1})y^{\alpha}]^{\vartheta+1}}, 0<y<1, \quad (6)$$

where $\alpha > 0$ and $\vartheta > 0$ are shape parameters, $0 < \tau < 1$ is the quantile parameter and $\rho \in [0,1]$. Figure 1 shows the PDF plot of the re-parameterized UG/G distribution for some selected parameter and quantile values. The PDF exhibits approximately symmetric, left skewed, right skewed, decreasing, increasing and bathtub shapes as shown in Figure 1 for the chosen parameter and quantile values.

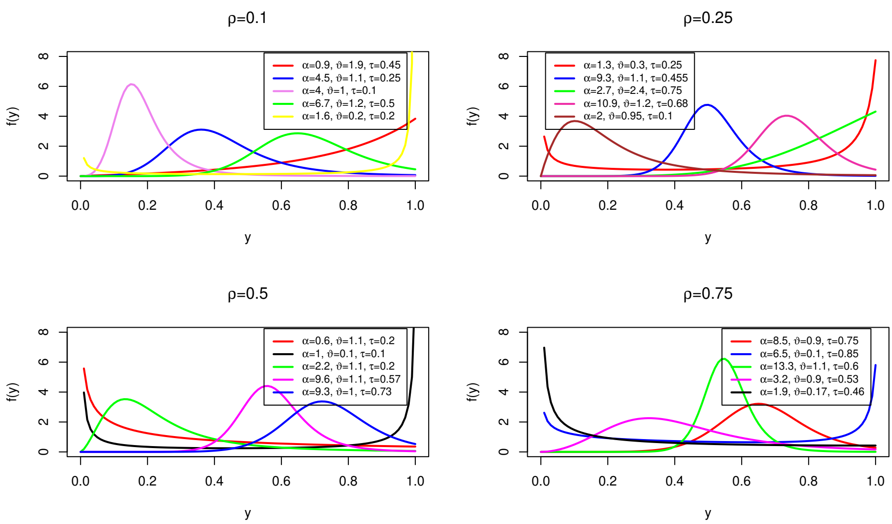

Fig. 1: **PDF plot of RUG/G for some selected parameter and quantile values**

The corresponding CDF of the RUG/G distribution is obtained by substituting θ into equation (2). The re-parameterized CDF is given by

$$F_{UG/G}(y;\alpha,\vartheta,\rho,\tau) = \frac{(\tau^{-\alpha}-1)^{\vartheta}}{(\tau^{-\alpha}-\rho^{-\frac{1}{\vartheta}}+(\rho^{-\frac{1}{\vartheta}}-1)y^{\alpha})^{\vartheta}}, 0<y<1. \quad (7)$$

The survival function, often known as the complimentary CDF is a function that estimates the likelihood that a patient, device, or other object of interest

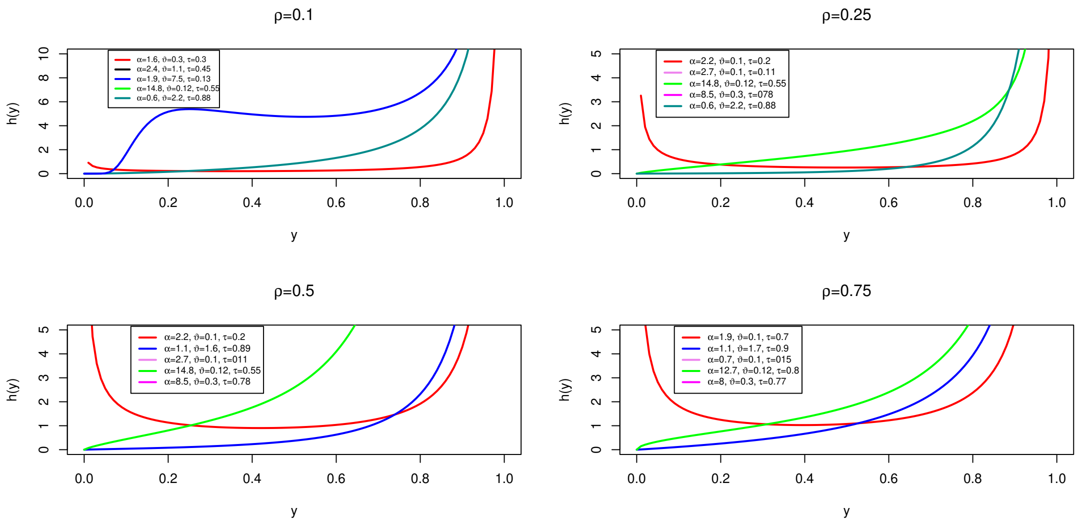

Fig. 2: Hazard rate function plot of the RUG/G distribution

will survive for a certain period of time. The survival function is sometimes referred to as the reliability function in the realm of engineering. The survival function is widely employed in a variety of domains, including applied biology, engineering, social and environmental sciences among others. It can be used to study the survival time of patients in medication trials. It may also be used to calculate the amount of time between employment transfers (Kleinbaun, 2012). The survival function of the RUG/G distribution is given by

$$S_{UG/G}(y) = \frac{(\tau^{-\alpha} - \rho^{-\frac{1}{\vartheta}} + (\rho^{-\frac{1}{\vartheta}} - 1)y^{\alpha})^{\vartheta} - (\tau^{-\alpha} - 1)^{\vartheta}}{(\tau^{-\alpha} - \rho^{-\frac{1}{\vartheta}} + (\rho^{-\frac{1}{\vartheta}} - 1)y^{\alpha})^{\vartheta}}, 0 < y < 1. \quad (8)$$

The hazard rate function of the re-parameterized UG/G distribution is given by

$$h_{UG/G}(y) = \frac{\alpha\vartheta(\rho^{\frac{-1}{\vartheta}} - 1)(\tau^{-\alpha} - 1)^{\vartheta}[(\rho^{\frac{-1}{\vartheta}} - 1) + (\tau^{-\alpha} - \rho^{-\frac{1}{\vartheta}} - 1)y^{\alpha}]^{-1}}{y^{1-\alpha\vartheta}([(\rho^{-\frac{1}{\vartheta}} - 1) + (\tau^{-\alpha} - \rho^{-\frac{1}{\vartheta}})y^{\alpha}]^{\vartheta} - (\tau^{-\alpha} - 1)^{\vartheta}y^{\alpha\vartheta})}, 0 < y < 1. \quad (9)$$

The plots of the hazard rate function for some selected parameter and quantile values are shown in Figure 2. The plots indicate that the hazard function of the RUG/G distribution has increasing shape and bath tub shape.

3 Results

3.1 The Unit Gamma/Gompertz quantile Regression Model

Let $Y_1, ..., Y_n$ be independent random variables such that each Y_i has the PDF described in equation (1), $Y_i \sim RUG/G(\tau_i; \alpha, \vartheta, \rho)$ for a known $\rho \in (0, 1)$ with quantile parameter τ_i and shape parameters $\alpha > 0$ and $\vartheta > 0$. The UG/G quantile regression model is expressed such that the quantile τ_i meets the following functional relation:

$$g(\tau_i) = \phi^T Z_i, i = 1, ..., n \tag{10}$$

where $\phi = (\phi_0, ..., \phi_{p-1})$ is a p-dimensional vector of unknown regression coefficients $(p < n)$ and $Z_i = (1, z_{i1,}..., z_{i(p-1)})$ represents the observations on p known covariates.

Since the quantile is between 0 and 1, the most useful well-known link functions for $g(.)$ are logit given by $logit(\tau_i) = log(\frac{\tau_i}{1-\tau_i})$, probit given by $g(\tau_i) = \Psi^{-1}(\tau_i)$, where $\Psi^{-1}(\tau_i)$ is the standard normal quantile function and complementary log-log given by $log(\tau_i) = log[-log(1 - \tau_i)]$. We consider the logit link in this study as a result of the direct explanation of the parameters. Thus we have

$$\frac{\tau_i}{1 - \tau_i} = exp(\phi_0 + \phi_1 z_{1i} + ... + \phi_p z_{pi}), \tag{11}$$

which leads to the following interpretations:

- If z_{1i} is continuous, for a unit increase, the percentage change in the mode response is $100\% \cdot (e^{\hat{\phi_1}} - 1)$, keeping the other predictors fixed.
- If z_{1i} is an indicator variable then $100\% \cdot e^{\hat{\phi_1}}$ represents the percentage change in the mode response for $z_{1i} = 1$ to $z_{1i} = 0$, holding the other predictors fixed.

3.2 Parameter Estimation of the UG/G quantile Regression Model

In this study, the MLE is employed to estimate the parameters of the UG/G regression model. Suppose $y_1, ..., y_n$ are random samples for the UG/G distribution with PDF defined by equation (6). Let $\xi = (\vartheta, \alpha, \tau_i)^T$ be the set of parameters of the distribution. From the likelihood function $L(\xi) = \prod_{i=1}^{n} f_{UG/G}(y_i; \xi)$ the total log-likelihood function is

$$\begin{aligned} \ell =& nlog(\alpha\vartheta) + \vartheta \sum_{i=1}^{n} log(\tau_i^{-\alpha} - 1) + (\alpha\vartheta - 1) \sum_{i=1}^{n} log(y_i) - \\ & nlog(\rho^{-\frac{1}{\vartheta}} - 1)^{\vartheta} - (\vartheta + 1) \sum_{i=1}^{n} [1 + (\frac{\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}}}{\rho^{-\frac{1}{\vartheta}} - 1})]. \end{aligned} \tag{12}$$

The maximum likelihood estimates of the parameter set $\hat{\xi} = (\hat{\vartheta}, \hat{\alpha}, \hat{\tau}_i)$ is obtain by differentiating equation (12) with respect to each parameter, equating it to zero and solving the system of equations for the parameters. The partial differentials are given by the following equations

$$\begin{aligned}\frac{\partial \ell}{\partial \vartheta} =& \frac{n}{\vartheta} - n[\frac{\rho^{-\frac{1}{\vartheta}} log(\rho)}{\vartheta(\rho^{-\frac{1}{\vartheta}} - 1)} + log(\rho^{-\frac{1}{\vartheta}} - 1)] + \alpha \sum_{i=1}^{n} log(y_i) + \\ & (\vartheta + 1) \sum_{i=1}^{n} \frac{\rho^{-\frac{1}{\vartheta}} log(\rho)[1 + (\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}})(\rho^{-\frac{1}{\vartheta}} - 1)^{-1}] y_i^{\alpha}}{\vartheta^2 [(\rho^{-\frac{1}{\vartheta}} - 1) + (\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}}) y_i^{\alpha}]} + \\ & \sum_{i=1}^{n} log(\tau_i^{-\alpha} - 1) - \sum_{i=1}^{n} [1 + (\frac{\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}}}{\rho^{-\frac{1}{\vartheta}} - 1}) y_i^{\alpha}],\end{aligned} \tag{13}$$

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} =& \frac{n}{\alpha} + (\vartheta + 1) \sum_{i=1}^{n} \frac{[\tau_i^{-\alpha} log(\tau_i) - (\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}}) log(y_i)] y_i^{\alpha}}{(\rho^{-\frac{1}{\vartheta}} - 1) + (\tau_i^{-\alpha} - \rho^{-\frac{1}{\vartheta}}) y_i^{\alpha}} - \\ & \vartheta \sum_{i=1}^{n} log(y_i) - \vartheta \sum_{i=1}^{n} \frac{\tau_i^{-\alpha} log(\tau_i)}{\tau_i^{-\alpha} - 1},\end{aligned} \tag{14}$$

and

$$\frac{\partial \ell}{\partial \phi_r} = \frac{\alpha(\vartheta + 1)}{(\rho^{-\frac{1}{\vartheta}} - 1)} \sum_{i=1}^{n} (\tau_i^{-(\alpha+1)} y_i^{\alpha}) \frac{\partial}{\partial \phi_r} \tau_i - \alpha\vartheta \sum_{i=1}^{n} \frac{\tau_i^{-(\alpha+1)}}{(\tau_i^{-\alpha} - 1)} \frac{\partial}{\partial \phi_r} \tau_i, \tag{15}$$

for $r = 1, ..., p$. The derivative $\frac{\partial}{\partial \phi_r} \tau_i$, will depend on the chosen link function. For example, if it is considered the logit link, which is given by

$$\tau_i = \frac{exp(Z_i^t \phi)}{1 + exp(Z_i^t \phi)}, \tag{16}$$

then the derivative in terms of the quantiles is given by

$$\frac{\partial}{\partial \phi_r} \tau_i = \tau_i (1 - \tau_i) z_{ir}, i = 1, ..., n, r = 1, ..., p. \tag{17}$$

3.3 Cox and Snell Residuals Analysis

The differences between the observed and predicted values are referred to as the residuals. They are the popular tool for determining a model's adequacy. If a model is adequate, the Cox–Snell residuals should be approximated to a unit exponential distribution, and the Cox–Snell residuals plotted against the cumulative hazard need to be a straight line with a zero intercept and unit gradient (Nasiru, 2021). The residuals of Cox-Snell are defined as

$$r = -log\hat{S}(y_i|x_i) = \hat{H}(y_i|x_i), \tag{18}$$

where $\hat{H}$ is the estimated cumulative hazard function based on the model, $\hat{s}(\cdot)$ is the estimated survival function and $\hat{S}(r_i)$ is the estimate of Kaplan–Meier of the Cox–Snell residuals.

4 Simulation Study

Monte Carlo simulations were used to evaluate the estimators for the parameters of the UG/G regression model. The two covariates were generated from uniform distribution for sample sizes $n = 30, 80, 130, 180$ and 230. The experiment was carried out 3000 times each one of the sample sizes. The following true parameter sets of values $I : (\phi_0, \phi_1, \phi_2, \alpha, \vartheta) = (0.8, 0.9, 1.4, 0.2, 5.0)$, $II : (\phi_0, \phi_1, \phi_2, \alpha, \vartheta) = (1.29, 0.9, 1.5, 0.2, 1.6)$ and $III : (\phi_0, \phi_1, \phi_2, \alpha, \vartheta) = (0.4, 0.5, 1.5, 0.6, 0.8)$ were used at $0.25, 0.50$, and 0.75 quantiles respectively to generate random samples from the UG/G model. Table 1 and 2 show the average value (AV), absolute bias (AB) and root mean square error (RMSE) for the estimators. As seen in Table 1, the AV varies with sample size and approaches the real parameter values as sample size increases. The ABs and RMSEs for all the parameters are positive, and reduces as the sample size grows. This shows that the consistency of the MLEs will be achieved as $n \to \infty$.

Table 1: Simulation Results for Unit Gamma/Gompertz Quantile Regression

PARAMETER	n	I AV	I AB	I RMSE	II AV	II AB	II RMSE	III AV	III AB	III RMSE
ϕ_0	30	-2.5624	3.2774	3.5785	-1.3262	2.8353	3.0964	-0.102	1.142	1.4457
	80	-2.5003	3.2003	3.3055	-1.3526	2.8526	2.9415	-0.0576	0.6826	0.8411
	130	-2.5022	3.2022	3.2653	-1.3475	2.8475	2.9026	-0.0407	0.5741	0.7078
	180	-2.5155	3.2005	3.2595	-1.3546	2.8456	2.8938	-0.0337	0.5309	0.6438
	230	-2.5245	3.1245	3.2507	-1.332	2.832	2.8605	-0.0289	0.4949	0.5896
ϕ_1	30	0.0552	1.6093	2.048	-0.0502	1.4839	1.8611	-0.0191	1.4666	1.8969
	80	-0.0207	1.083	1.3368	-0.0028	0.9534	1.1744	0.0151	0.8558	1.0689
	130	-0.0024	0.9388	1.1546	-0.0181	0.8356	1.0222	-0.0036	0.7153	0.8909
	180	0.0122	0.8779	1.0556	0.0041	0.7795	0.932	0.0107	0.6339	0.7829
	230	0.0103	0.8547	1.0103	-0.008	0.7519	0.8863	-0.0076	0.607	0.7461
ϕ_2	30	-0.0067	1.8748	2.3395	-0.0226	1.6567	2.0356	-0.0002	1.8909	2.3407
	80	-0.0111	1.4171	1.6869	-0.0079	1.3013	1.5317	-0.0167	1.5517	1.7823
	130	-0.0111	1.3595	1.5522	0.0179	1.2198	1.406	-0.002	1.515	1.6698
	180	0.0015	1.3119	1.4723	0.0127	1.2024	1.3499	-0.014	1.5143	1.6315
	230	0.0034	1.305	1.4378	-0.0116	1.2173	1.3297	0.0138	1.4874	1.5787
α	30	2.962	2.7765	5.5601	2.1302	1.9538	5.8894	11.7228	11.501	57.1874
	80	2.2755	2.1142	5.3283	1.9129	1.7825	4.7444	3.0378	2.8496	9.3239
	130	2.0398	1.9021	4.4283	1.8602	1.7379	4.6024	3.0343	2.8416	8.886
	180	1.744	1.6102	3.6178	1.6952	1.5838	3.7753	2.8001	2.6007	7.6187
	230	1.7268	1.6036	3.4303	1.5692	1.4672	3.4818	2.7526	2.5557	7.3986
ϑ	30	72.7297	72.1899	103.949	62.7267	62.2806	93.287	49.6607	49.365	80.3951
	80	58.1775	57.8708	97.1921	54.7544	54.3873	89.8144	41.2763	41.0181	74.0449
	130	55.5244	55.7652	95.6587	51.8524	51.4797	87.6113	39.8738	39.6163	72.7697
	180	56.141	55.1413	96.4809	51.6195	51.2338	87.5141	36.2566	36.014	69.1556
	230	56.7788	55.3532	97.4531	51.3758	50.9758	87.2882	34.615	34.3617	67.3595

5 Applications of the UG/G quantile regression model Using Real-life Data

In this section, we utilize three real life datasets to show the efficacy of the UG/G quantile regression model. In the first example, we used the Gasoline Yields from Crude Oil data present in the betareg package in R. In the second example, the long term interest (LTI) rates against the foreign direct investment (FDI) data found in Nasiru et al. (2021) and Altun et al. (2020) is used. While in the third example, the ammonia data available in the ugomquantreg R package is utilized. The logit link function was considered in all cases. The performance of the UG/G quantile regression model was compared with the classical beta regression model using goodness of fit measure such us the -2log-likelihood function (-2ℓ) and the Akaike Information Criteria (AIC). In order to model different locations within the distribution of the response variable and also to investigate how the changes in the covariates affect the behavior of the response variable,$0.200, 0.500, 0.800, 0.900$ and 0.999 quantile values were selected for the quantile model specifications. Again the 0.999 quantile value was employed to measure an extreme value impact, which is unlikely to be captured when using the mean regression.

5.1 Gasoline Yields from Crude Oil Data

The dataset is made up of 6 variables containing 32 observations on the response and 32 observations on the independent variables. The dataset was collected by Prater (1956). Proportion of crude oil converted to gasoline after distillation and fractionation (yield) is the response variable, while crude oil gravity (gravity), pressure vapor pressure of crude oil (pressure), temperature at which 10 percent of crude oil has vaporized (temp10), temperature at which all gasoline has vaporized (temp) and factor indicating unique batch of conditions gravity, pressure, and temp10 (batch) are the independent variables. In this study all the variables were considered except the batch.

The descriptive statistics of the data shown in Table 2 revealed that the dependent variable has the minimum (min) and maximum (max) values of 0.0280 and 0.457 respectively. The mean yield is 0.1966. The coefficient of skewness (skew) is 0.3687 and excess-kurtosis value is -0.8003 indicated that the data was right skewed and less peaked compared to the normal distribution. The coefficients of skewness of the covariates revealed that the covariates are right skewed except the temp data which is negatively skewed.

Table 3 shows the parameter estimates, standard errors (SE), and p-value for five quantile functions, $\rho_i = (0.200, 0.500, 0.800, 0.900, 0.999)$ and the beta model. The estimates of pressure parameters are not significant at 5% except

Table 2: Descriptive statistics of the GasonlineYield data

	yield	**gravity**	**pressure**	**temp10**	**temp**
Mean	0.1966	39.2500	4.1813	241.5000	332.0938
Sd	0.1072	5.6354	2.6198	37.5414	69.7560
Min	0.0280	31.8000	0.2000	190.0000	205.0000
Max	0.4570	50.8000	8.6000	316.0000	444.0000
Skew	0.3687	0.5612	0.1101	0.4934	-0.2658
Kurtosis	-0.8003	-0.1222	-1.1667	-0.8439	-1.1941

at the 0.900 quantile. Also the estimates of gravity parameters are not significant at 0.8, 0.9, 0.999 and in the beta model at 5%.

Table 3: The maximum likelihood parameter estimates for the GasonlineYield datasets

Quantile		**Parameter**						
ρ		**intercept**	**gravity**	**pressure**	**temp10**	**temp**	α	ϑ
0.2	Estimate	1.3832	-0.0241	-0.0515	-0.0245	0.0117	6.2192	10.5573
	SE	-0.5001	0.0093	0.0263	0.0016	0.0006	0.4658	0.0150
	p-value	0.0057	0.0091	0.0500	<0.0001	<0.0001	<0.0001	<0.0001
0.5	Estimate	0.6364	-0.0266	-0.0238	-0.0214	0.0120	9.4307	10.4220
	SE	0.0004	0.0059	0.0185	0.0010	0.0005	0.0001	0.0000
	p-value	<0.0001	<0.0001	0.1981	<0.0001	<0.0001	<0.0001	<0.0001
0.8	Estimate	-1.0560	-0.0091	0.0007	-0.0175	0.0125	9.8083	10.4730
	SE	0.0004	0.0054	0.0167	0.0009	0.0004	0.0000	0.0000
	p-value	<0.0001	0.0917	0.9682	<0.0001	<0.0001	<0.0001	<0.0001
0.9	Estimate	-2.3912	0.0010	0.0533	-0.0138	0.0128	8.2871	1.4511
	SE	0.0088	0.0076	0.0262	0.0013	0.0007	0.1457	0.4946
	p-value	<0.0001	0.8937	0.0422	<0.0001	<0.0001	<0.0001	<0.0001
0.999	Estimate	-1.2201	0.0009	0.0258	-0.0197	0.0160	9.4243	2.5250
	SE	0.0006	0.0075	0.0251	0.0014	0.0007	0.0000	0.0001
	p-value	<0.0001	0.9024	0.3034	<0.0001	<0.0001	<0.0001	<0.0001
Beta	Estimate	-2.6949	0.0045	0.0304	-0.0110	0.0106		248.2400
	SE	0.7626	0.0071	0.0281	0.0026	0.0005		62.0200
	p-value	0.0004	0.5249	0.2791	<0.0001	<0.0001		0.0001

The goodness-of-fit statistics in the case of fitted models are shown in Table 4. The UG/G quantile regression model at the 0.999 quantile is the most appropriate model for the data since it has the lowest -2ℓ and AIC values. The model is shown in equation (19) and it could be seen that gravity, pressure and temp have positive influence on the yield whiles the temp10 has negative influence on the yield.

$$\begin{aligned} yield = & -1.2201 + 0.0009gravity + 0.0258pressure- \\ & 0.0197temp10 + 0.0160temp \end{aligned} \tag{19}$$

Table 4: Goodness-of-fit statistics for GasonlineYield dataset

Quantile	-2ℓ	**AIC**
0.2000	-118.0121	-104.0121
0.5000	-116.7779	-102.7779
0.8000	-135.0878	-121.0878
0.9000	-144.6250	-130.6250
0.9990	-154.9649*	-140.9649*
Beta	-151.3600	-139.3614

*: means best fit model

In order to assess the fitted models, the Cox Snell residuals were generated. If the model is appropriate, then the Cox-Snell residuals should behave approximately in the same way as a sample from the standard exponential distribution (Lawless, 2003). Figure 3 revealed that probability-probability (P-P) plot, where the empirical probabilities of the Cox Snell residuals are compared with those of the standard exponential distribution. The depicted points for the UG/G quantile regression at the 0.999 quantile are noted to be closer to the diagonal line than those for the traditional beta regression. This implies that the UG/G quantile regression provides better fit than the classical beta regression.

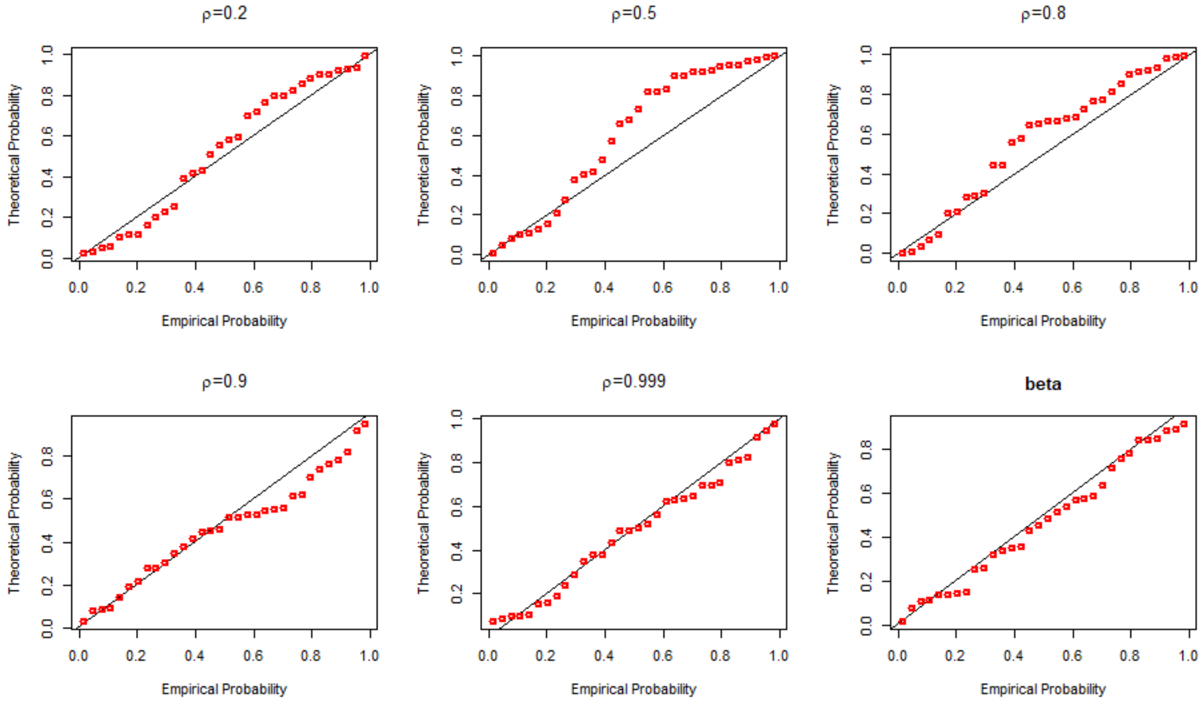

Fig. 3: Cox–Snell residuals plots for the UG/G quantile regression model

For model diagnostics, we have taken into account the goodness-of-fit measures such as the Kolmogorov-Smirnov statistic (KS), the Cramér-von Mises statistic (CVM) and the Anderson-Darling statistic (AD) of the Cox Snell residuals. The Cox-Snell residuals are said to follow an exponential distribution of parameter 1 if the fitted model is good. The statistics with smaller values and larger p-values are said to be better fits.
Table 5 shows the model diagnotics results for the GasonlineYield data. Table 5 revealed that the data sets can be modeled using the UG/G quantile regression model and the classical beta regression except at the 0.5 quantile. In terms of goodness of fit statistics, the UG/G quantile regression model at the 0.999 quantile outperforms the UG/G quantile regression at other selected quantiles and the classical beta for the data sets.

Table 5: Residuals goodness-of-fit statistics for GasonlineYield data

	KS		**CVM**		**AD**	
Quantile	**Statistic**	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**
0.200	0.1397	0.5153	0.1504	0.3898	0.9434	0.3876
0.500	0.2875	0.0078	0.6452	0.0167	4.6193	0.0044
0.800	0.2104	0.1013	0.2664	0.1686	1.6305	0.1484
0.900	0.1623	0.3320	0.1513	0.3872	0.8443	0.4492
0.999	0.1057	0.8306	0.0444	0.9122	0.3459	0.8994
Beta	0.1034	0.8488	0.0644	0.7896	0.4797	0.7663

5.2 Long Term Interest rates against Foreign Direct Investment data

The application of UG/G quantile regression is demonstrated by simulating the link between the Organization for Economic Cooperation and Development (OECD) countries' long-term interest rates and foreign direct investment (FDI). From Table 6, it could be observed that the averages of the LTI and the FDI were 0.0200 and 0.6469 with standard deviations of 0.0200 and 0.8848 respectively. The degrees of asymmetry (Skew) of the distributions of the two data sets were 1.4268 and 2.2861 respectively. These imply that the distributions are skewed to the right with a heavier right tail. The extents to which the data fall within the centre of the distribution were 2.1029 and 4.4225 which indicate that the distributions were leptokurtic with higher peaks as compared to the standard normal distribution.

Table 7 gives the MLEs, their standard errors (SEs) and corresponding p-values for the UG/G quantile and the classical beta regressions fitted to the

Table 6: Descriptive statistics of the LTI FDI datasets

statistics	LTI	FDI
mean	0.0200	0.6469
Sd	0.0200	0.8848
Min	0.0004	0.0356
Max	0.0775	3.8010
Skew	1.4268	2.2861
kurtosis	2.1029	4.4225

LTI rates. The p-values in Table 8 reveal that the parameters are statistically significant at 5% level for both regression models. We conclude that the FDI stocks explain the LTI rates.

Table 7: The maximum likelihood parameter estimates for LTI FDI data

Quantile		Parameter			
ρ		intercept	FDI	α	ϑ
0.2	Estimate	-4.8344	-0.5234	4.5419	0.2087
	SE	0.3033	0.1076	1.7270	0.1041
	p-value	<0.0001	<0.0001	0.0085	0.0451
0.5	Estimate	-3.8479	-0.5261	4.5398	0.2087
	SE	0.1802	0.1081	1.7254	0.1042
	p-value	<0.0001	<0.0001	0.0085	0.0450
0.8	Estimate	-3.2522	-0.5296	4.5373	0.2089
	SE	0.1586	0.1087	1.7231	0.1041
	p-value	<0.0001	<0.0001	0.0085	0.0448
0.9	Estimate	-3.0073	-0.5317	4.5379	0.2088
	SE	0.1623	0.1090	1.7224	0.1040
	p-value	<0.0001	<0.0001	0.0084	0.0447
0.999	Estimate	-1.8462	-0.5530	4.5867	0.2060
	SE	0.4884	0.1093	1.7191	0.1013
	p-value	0.0002	0.0000	0.0076	0.0421
Beta	Estimate	-3.7591	-0.3716		71.2100
	SE	0.1686	0.1722		18.8100
	p-value	<0.0001	0.0309		0.0002

The performance of the UG/G quantile regression model was compared to that of the beta regression model. The -2ℓ and AIC values are used to select the best fitted regression. Since the UG/G quantile regression has the lowest values of these statistics, it provides a better fit than the classical beta regression for the current data as shown in Table 8.
The best UG/G quantile regression model occurred at the 0.20 quantile value. Its mathematical expression is shown in equation (20). It could be observed that the FDI has negative influence on the LTI. This means that one unit increase in FDI will reduce the LTI by 52.34%.

$$LTI = -4.8344 - 0.5234FDI \tag{20}$$

Table 8: Goodness-of-fit statistics for LTI FDI data

Quantile	-2ℓ	**AIC**
0.2	-211.6940*	-203.6940*
0.5	-211.6783	-203.6783
0.8	-211.6581	-203.6581
0.9	-211.6457	-203.6457
0.999	-211.5246	-203.5246
Beta	-207.6	-201.5314

*: means best fit model

The plots of Cox–Snell residuals for the UG/Gquantile and the classical beta regressions are displayed in Figure 4. The P-P plots of the Cox–Snell residuals revealed that the UG/G quantile regression at the various quantile functions provide adequate fit to these data.

ISSN-2424-6271

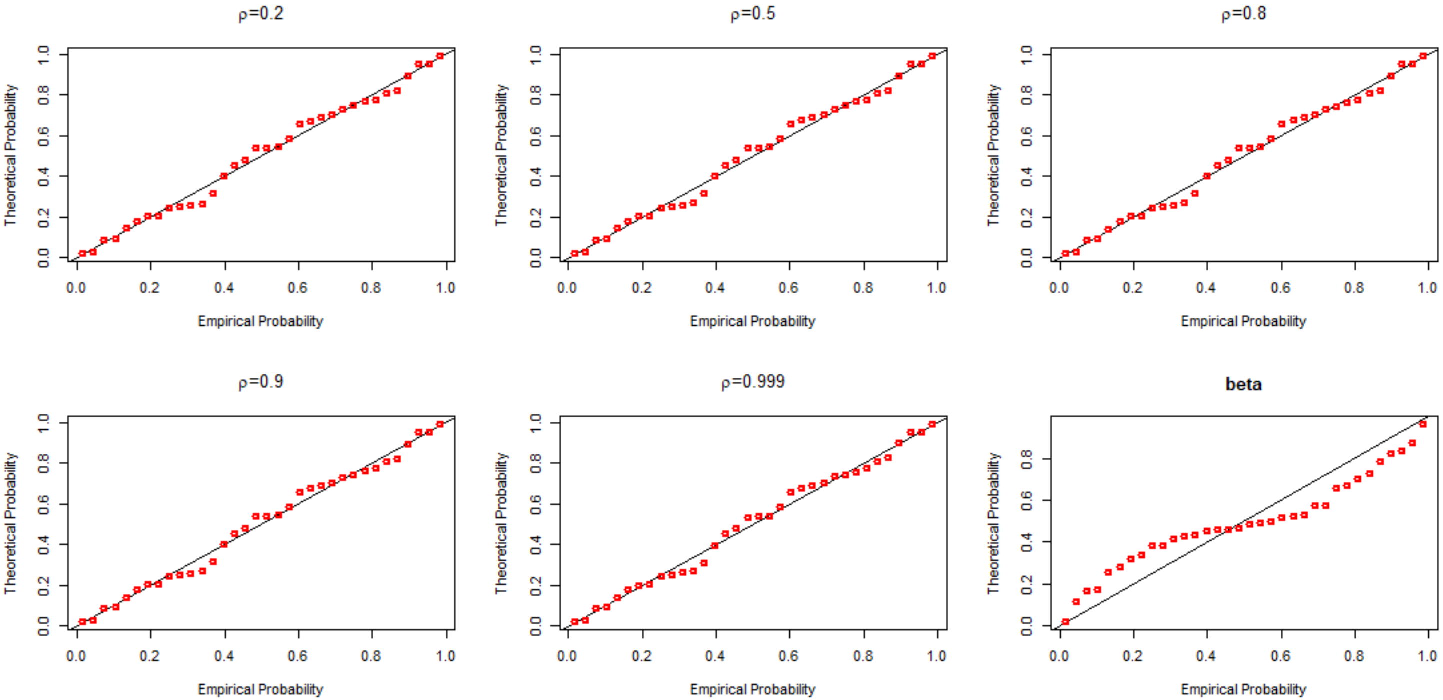

Fig. 4: Cox–Snell residuals plots for the UG/G quantile regression model

Table 9 shows the model diagnosis results for the LTI FDI data. Table 9 revealed that the data sets can be modeled using the UG/G quantile regression model and the classical beta regression. In terms of goodness-of-fit statistics, the UG/G quantile regression model performs better than the classical beta for the data sets.

Table 9: Residuals goodness-of-fit statistics for LTI FDI data

	KS		CVM		AD	
Quantile	**Statistic**	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**
0.20	0.0866	0.9415	0.0303	0.9769	0.1903	0.9929
0.50	0.0863	0.9427	0.0304	0.9766	0.1906	0.9928
0.80	0.0861	0.9440	0.0305	0.9762	0.1912	0.9927
0.90	0.0859	0.9447	0.0305	0.9760	0.1915	0.9926
1.00	0.0845	0.9515	0.0309	0.9747	0.1938	0.9921
Beta	0.1608	0.3086	0.2896	0.1445	1.5899	0.1566

5.3 Ammonia Data

The data are from the operation of a plant where ammonia is oxidized to nitric acid; the oxidation is measured on 21 consecutive days. The data set consist of 3 covariates and 1 response variable. The percentage of ammonia lost (StackLoss) is the response variable while the air flow to the plant (AirFlow), the cooling water inlet temperature (WaterTemp) and the acid concentration (AcidConc) are the covariates. Table 10 presents the descriptive statistics of the Ammonia data. From Table 10, the StackLoss has minimum and maximum values of 0.0700 and 0.4200 respectively. The average was 0.1752 with a standard deviation of 0.1017. The coefficient of skewness of 1.1564 and excess kurtosis value of 0.1344 revealed that the distribution of StackLoss is right-skewed, and it is highly peaked than the normal distribution due its value of excess kurtosis.

The parameter estimations are provided in Table 11. The parameter estimates were significant at the 5% significance level for all the independent variables except the AcidConc, which has its coefficients fluctuating between positive and negative and have p-values larger than 5%. This means that, the covariate, AcidConc has no influence on the response variable.

Table 12 presents the goodness-of-fit statistics for the fitted models. It can be

Table 10: Descriptive statistics of Ammonia data

Statistics	**StackLoss**	**AirFlow**	**WaterTemp**	**AcidConc**
Mean	0.1752	60.4286	21.0952	86.2857
Sd	0.1017	9.1683	3.1608	5.3586
Min	0.07	50	17	72
max	0.42	80	27	93
skew	1.1564	0.8119	0.4688	-0.873
kurtosis	0.1344	-0.2592	-1.2315	0.1896

Table 11: The maximum likelihood parameter estimates for Ammonia data

Quantile		Parameter					
ρ		intercept	AirFlow	WaterTem	AcidConc	α	ϑ
0.2	Estimate	-6.1169	0.0478	0.0686	-0.0005	10.6418	1.5466
	SE	0.5017	0.0099	0.0254	0.0077	0.4039	0.7948
	p-value	<0.0001	<0.0001	0.0070	0.9506	<0.0001	0.0517
0.5	Estimate	-6.2439	0.0514	0.0628	0.0015	11.4249	1.2354
	SE	0.4896	0.0101	0.0245	0.0078	0.1655	0.5899
	p-value	<0.0001	<0.0001	0.0104	0.8488	<0.0001	0.0363
0.8	Estimate	-5.7050	0.0461	0.0909	-0.0061	8.6632	3.4454
	SE	0.5934	0.0082	0.0225	0.0081	1.4573	0.4848
	p-value	<0.0001	<0.0001	0.0001	0.4552	<0.0001	<0.0001
0.9	Estimate	-5.9320	0.0519	0.0796	-0.0034	9.4776	2.0504
	SE	0.5738	0.0107	0.0261	0.0084	0.8384	0.8754
	p-value	<0.0001	<0.0001	0.0023	0.6830	<0.0001	0.0192
0.999	Estimate	-6.6419	0.0706	0.0616	0.0017	12.7830	0.9018
	SE	0.5797	0.0105	0.0258	0.0080	3.1016	0.4773
	p-value	<0.0001	<0.0001	0.0168	0.8350	0.0000	0.0588
beta	Estimate	-5.8466	0.0394	0.0905	-0.0012		207.0600
	SE	0.7727	0.0075	0.0218	0.0099		63.8600
	p-value	<0.0001	<0.0001	<0.0001	0.9010		0.0012

seen from Table 12 that the UG/G quantile regression models for the selected quantile values performed better than the classical beta regression model since they have the least -2ℓ and AIC values.
The UG/G quantile regression model at the 0.999 quantile value is the best among the selected quantiles since it records the least of the goodness-of-fit-measures. Its expression is shown in equation (21.The AirFlow and Wa-

terTemp have a positive effect on the StackLoss, while the AcidConc has a negligible effect.

$$\begin{aligned} StackLoss =& 6.6419 + 0.0706 AirFlow + 0.0616 WaterTem + \\ & 0.0017 AcidConc \end{aligned} \tag{21}$$

Table 12: Goodness-of-fit statistics for Ammonia data

Quantile	-2ℓ	**AIC**
0.200	-99.3219	-87.3219
0.500	-99.8405	-87.8405
0.800	-98.5533	-86.5448
0.900	-99.6211	-87.6211
0.999	-101.7153*	-89.7153*
Beta	-96.2200	-86.2127

*: means best fit model

Figure 5 reports the P-P plots of the Cox Snell residuals of UG/G quantile regressions and the classical beta regression. It is observed that the plotted points for the UG/G quantile regression at 0.999 quantile are closer to the diagonal line than those of the classical beta regression. This implies that the UG/G quantile regression provides better fit than the classical beta regression.

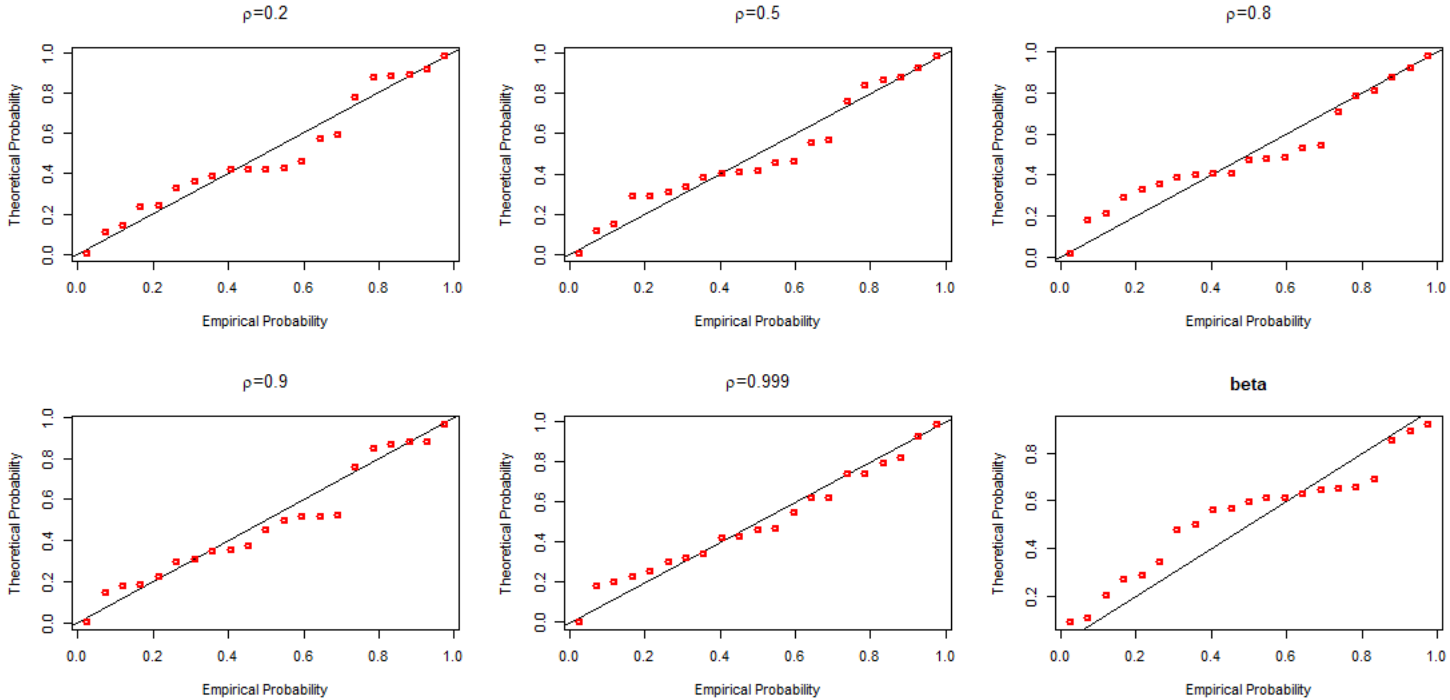

Fig. 5: Cox–Snell residuals plots for the UG/G quantile regression model

Table 13 shows the model diagnosis results for the Ammonia data. Table 13 revealed that the data sets can be modeled using the UG/G quantile regression model and the classical beta regression. In terms of goodness-of-fit statistics, the UG/G quantile regression model performs better than the classical beta for the data sets.

Table 13: Residuals goodness-of-fit statistics for Ammonia data

	KS		**CVM**		**AD**	
Quantile	**Statistic**	**P-Value**	**Statistic**	**P-Value**	**Statistic**	**P-Value**
0.200	0.1579	0.6162	0.0877	0.6536	0.4897	0.7554
0.500	0.1540	0.6461	0.0929	0.6253	0.5061	0.7386
0.800	0.1706	0.5192	0.1259	0.4753	0.6594	0.5917
0.900	0.1911	0.3787	0.0855	0.6655	0.4956	0.7493
0.999	0.1345	0.7945	0.0543	0.8552	0.4117	0.8355
Beta	0.1909	0.1909	0.1908	0.2874	1.0125	0.3498

6 Conclusion

In this study, a new unit quantile regression model was developed. The proposed model is based on a re-parameterization of the UG/G distribution in terms of its quantiles. MLEs was employed for estimating the parameters of the regression model. A Monte Carlo simulation study was performed and has shown that the parameters were well estimated in terms of the bias and RMSE. Three real life datasets were analyzed for illustrative and model comparison purposes. For these datasets, the UG/G quantile regression model has outperformed the beta models according to the information criteria, goodness-of-fit measures and the P-P plots for the Cox–Snell residuals. Subsequent study in this area should focus on the modal regression to determine whether it will produce a better fit than the classical beta model.

References

Altun, E. (2021) The log-weighted exponential regression model: alternative to the beta regression model, *Communications in Statistics-Theory and Methods*, 50(10): 2306-2321.

Altun, E., and Cordeiro, G. M. (2020) The unit-improved second-degree Lindley distribution: inference and regression modeling, *Computational Statistics*, 35(1): 259-279.

Altun, E., El-Morshedy, M., and Eliwa, M. S. (2021) A new regression model for bounded response variable: An alternative to the beta and unit-Lindley regression models, *Plos one*, 16(1): e0245627.

Bantan, R. A., Jamal, F., Chesneau, C., and Elgarhy, M. (2021) Theory and applications of the unit gamma/gompertz distribution, *Mathematics*, 9(16): 1850.

Barndorff-Nielsen, O. E., and Jørgensen, B. (1991) Some parametric models on the simplex, *Journal of multivariate analysis*, 39(1), 106-116.

Bayes C.L., Bazán J.L., and García C., (2012) A new robust regression model for proportions, *Bayesian Anal*, 7: 841–866.

Bayes, C. L., Bazán, J. L., and De Castro, M. (2017) A quantile parametric mixed regression model for bounded response variables, *Statistics and its interface*, 10(3): 483-493.

Bemmaor, A. C., and Glady, N. (2012) Modeling purchasing behavior with sudden "death": A flexible customer lifetime model, *Management Science*, 58(50): 1012-1021.

Bonat, W. H., Petterle, R. R., Hinde, J., and Demétrio, C. G. (2019) Flexible quasi-beta regression models for continuous bounded data, *Statistical Modelling*, 19(6): 617-633.

Gómez-Déniz, E., Sordo, M. A., and Calderín-Ojeda, E. (2014) The Log–Lindley distribution as an alternative to the beta regression model with applications in insurance, *Insurance: Mathematics and Economics*, 54: 49-57.

Jodra, P., and Jimenez-Gamero, M. D. (2020) A quantile regression model for bounded responses based on the exponential-geometric distribution, *Revstat*, 4: 415–436.

Korkmaz, M. Ç. (2020) The unit generalized half normal distribution: A new bounded distribution with inference and application, *UPB Scientific Bulletin*, 82: 133-140.

Korkmaz, M. Ç., Chesneau, C., and Korkmaz, Z. S. (2021) Transmuted unit Rayleigh quantile regression model: alternative to beta and Kumaraswamy quantile regression models, *Univ. Politeh. Buchar. Sci. Bull. Ser. Appl. Math. Phys*, 83: 149-158.

Korkmaz, M. Ç., Altun, E., Alizadeh, M., and El-Morshedy, M. (2021) The Log Exponential-Power Distribution: Properties, Estimations and Quantile Regression Model, *Mathematics*, 9(21): 2634.

Kumaraswamy, P. (1980) A generalized probability density function for double-bounded random processes, *Journal of hydrology*, 46(1-2): 79-88.

Lawless, J. F. (2003) *Basic concepts and models 1.1. statistical models and methods for lifetime data*, 1-47.

Lemonte A. J. and Bazán JL. (2016) New class of Johnson SB distributions and its associated regression model for rates and proportions, *Biometric J.*, 41: 727–746.

Mazucheli, J., Menezes, A. F. B., and Chakraborty, S. (2019) On the one parameter unit-Lindley distribution and its associated regression model for proportion data, *Journal of Applied Statistics*, 46(4): 700

Mazucheli, J., Menezes, A. F. B., Fernandes, L. B., de Oliveira, R. P., and Ghitany, M. E. (2020) The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates, *Journal of Applied Statistics*, 47(6): 954-974.

Mitnik, P. A., and Baek, S. (2013) The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation, *Statistical Papers*, 54(1): 177-192.

Mousa A. M., El-Sheikh A. A., and Abdel-Fattah M. A. (2016) A gamma regression for bounded continuous variables, *Advances and Applications in Statistics*, 49(4): 305–326.

Nasiru, S., Abubakari, A. G., and Angbing, I. D. (2021) Bounded Odd Inverse Pareto Exponential Distribution: Properties, Estimation, and Regression, *International Journal of Mathematics and Mathematical Sciences*, 2021, 1-18.

Paz, R. F. D. (2017) *Alternative regression models to beta distribution under bayesian approach*.

Prater, N.H. (1956) Estimate Gasoline Yields from Crudes, *Petroleum Refiner*, 35(5): 236–238.

Santos, B., and Bolfarine, H. (2015) Bayesian analysis for zero-or-one inflated proportion data using quantile regression, *Journal of Statistical Computation and Simulation*, 85(17): 3579-3593.