Dhaka Univ. J. Sci. 74(2): 361–366, 2026 (July)
ISSN 1022-2502, E-ISSN 2408-8528
DOI: https://doi.org/10.3329/dujs.v74i2.86968

Modeling Bounded Response Variables Using Quantile Regression Based on Arc Tangent Unit Weibull Distribution

Roosha and Md. Jamil Hasan Karami*

Department of Statistics, University of Dhaka, Dhaka–1000, Bangladesh

(Received: 7 January 2026; Accepted: 25 March 2026)

Abstract

This study introduces the Arc Tangent Unit Weibull (ATUW) distribution, a model for proportion based educational attainment data bounded within the unit interval. the ATUW model,which is derived by applying an arc tangent transformation to the classical Weibull distribution, provides greater flexibility in capturing skewed outcomes and extreme values. To estimate the parameters of the proposed model, several estimation methods are applied and their performance are compared through simulation studies, focusing on bias and mean squared error under varying conditions. The model is further reparametrized as the Quantile Arc Tangent Unit Weibull (QATUW) distribution, enabling its use in quantile regression analysis. Application to district-level data from the Bangladesh Multiple Indicator Cluster Survey (2019) reveals maternal education and age at first school enrollment as significant predictors of children's higher educational attainment. Comparative results demonstrate that the QATUW model outperforms traditional Beta and Kumaraswamy distributions, particularly in modeling extreme quantiles, offering stronger evidence for educational policy and planning.

Keywords: Quantile regression, bounded response variables, Arc Tangent Unit Weibull distribution (ATUW), re-parameterization, hyperbolic transformation, educational data.

I. Introduction

In a variety of fields, including reliability analysis, biostatistics, finance, and environmental research, bounded response variables those restricted to a predetermined range are frequently employed in statistical modeling. Skewness, long tails, and heteroscedasticity variations in spread over the response range are examples of non-normal characteristics that are frequently seen in these variables. The intricacy of such data may be lost in traditional mean regression techniques, which estimate the conditional mean of the response variable. This is particularly true when non-constant variance and departures from normality occur. This restriction emphasizes the importance of developing techniques that can model the entire conditional distribution of the response variable.

A robust and effective alternative that overcomes many of these restrictions is quantile regression, proposed by Koenker[1]. Quantile regression models the conditional quantiles of the response variable as opposed to the mean. This method allows for an expanded understanding of the relationship between predictors and response by enabling a comprehensive investigation of how predictors influence the response distribution at different quantile levels (such as the median, lower, and higher tails). Because it accounts for non-normality and variations across the entire range of the response, quantile regression is particularly advantageous for modeling bounded response variables.

A generalization of the Weibull distribution, the Arc Tangent Unit Weibull (ATUW) distribution, has been proposed recently as a versatile parametric distribution for modeling bounded, skewed, and asymmetric data. It is an effective option for modeling bounded response variables due to its shape parameters, which allow it to adapt to a wide range of data distributions. The application of this distribution to complex patterns in real-world data is enhanced by integrating it with quantile regression. Ferrari et al.[2] proposed the Beta distribution for cases where the response variable is bounded, but it can be inadequate for both modeling and prediction in real-world data settings. Rangvid[3] studied school composition effects in Denmark using quantile regression with data from the Program for International Student Assessment (PISA) 2000. Eide et al[4]. employed quantile regression to measure the effect of school quality on student performance. Buchinsky[5] assessed the dynamics of changes in the female wage distribution in the USA using the quantile regression approach. The Weibull distribution, introduced by Weibull[6], is a flexible statistical distribution that is particularly useful for modeling failure times, reliability analysis, and life data analysis.[6] Korkmaz et al[7]. introduced the Arc Secant Hyperbolic Normal distribution and explored its properties, quantile regression modeling, and applications.

By combining quantile regression with the recently established ATUW distribution, which is specifically designed to capture the unique characteristics of bounded data this study presents an approach for modeling bounded response variables. The proposed framework allows for precise modeling of conditional quantiles for bounded variables, providing enhanced flexibility and versatility.

* Author for correspondence. e-mail: karami.stat@du.ac.bd

The technique is applied to real-world datasets to validate its efficacy and demonstrate its ability to capture complex data behaviors. Educational data obtained from the Bangladesh Multiple Indicator Cluster Survey has been used for this analysis. Additionally, the paper illustrates the superior flexibility, robustness, and accuracy of the ATUW distribution by comparing it with other well-known bounded distributions, such as the Kumaraswamy, and Beta distributions, which were proposed by Kumaraswamy[8], and Ferrari et al[2]., respectively.

II. Data and Methods

For the analysis purpose, data have been extracted from the Bangladesh Multiple Indicator Cluster Survey (MICS). This is a nationally representative, cross-sectional household survey conducted by the Bangladesh Bureau of Statistics in collaboration with UNICEF. It employs a stratified two-stage sampling technique based on the national census and gathers high-quality, globally comparable data on women, households, and children[9], which collects data from more than 60,000 households, facilitates tracking progress against both national and international commitments, including the Sustainable Development Goals (SDGs), and enables national and subnational analysis.

The child's greatest degree of education, which ranges from no formal education to higher secondary and beyond, the proportion of children completed higher education from each district is the dependent variable in this study. The elements that followed are the independent variables which are taken as the proportion of mother's who has completed higher level education of each district, a crucial sociodemographic factor that is strongly linked to the educational attainment of the child; mean age at first school attendance, which indicates the child's entry into formal education and can affect long-term academic progression; and someone read books to child, that evaluates a child's capacity to read age-appropriate texts and indicates cognitive development and learning outcomes.

Methodology

We introduce a random variable T having the Weibull distribution with parameters α > 0 and β > 0, with the cumulative distribution function. i.e.,

$$F(t,\alpha,\beta) = 1 - e^{-\alpha t^{\beta}}.$$

We defined the transformed variable

$$X = \frac{2}{\pi}\arctan(T),$$

which takes values in the interval (0,1).

Here the cdf of X is as follows

$$F(x;\ \alpha,\beta) = 1 - exp\left[-\alpha\left(tan\left(\frac{\pi x}{2}\right)\right)^{\beta}\right], x \in (0,1)$$

Probability Density Function(pdf):

$$f(x;\alpha,\beta) = \alpha\beta(A)^{\beta-1}\exp\left[-\alpha(A)^{\beta}\right]\frac{\pi}{2}\sec^2\left(\frac{\pi x}{2}\right),$$

$$x\epsilon\ (0,1),$$

where, $A = \tan\frac{\pi x}{2}$.

Also got The percentiles of the ATUW distribution are defined by the inverse function of $F\ (x,\alpha,\beta)$. Further it will be used to get random number of simulation study.

$$x_u(\alpha,\beta) = \frac{2}{\pi}\arctan\left[\left(-\frac{1}{\alpha}\log(1-u)\right)^{1/\beta}\right].$$

The median is as given bellow

$$M = x_{0.5}(\alpha,\beta) = \frac{2}{\pi}\arctan\left[\left(\frac{log2}{\alpha}\right)^{\frac{1}{\beta}}\right].$$

A number of estimation techniques, which include Anderson–Darling Estimation (ADE), Cramér–von Mises Estimation (CVME), Least Squares Estimation (LSE), Weighted Least Squares Estimation (WLSE), Maximum Likelihood Estimation (MLE), and Maximum Product of Spacings Estimation (MPSE), have been employed to estimate the parameters of the proposed distribution.

Key measures such mean squared error (MSE), bias, and average estimates (AE) were used to evaluate the estimators. These measures investigate how closely and precisely the parameters are estimated.

Then we have carried out a univariate data analysis to see modeling ability of the ATUW distribution. Our current goal is to evaluate the ATUW distribution's adjustment power using the maximum likelihood approach with that of the unit distributions that are referenced, which include the beta, Kumaraswamy (Kw). By modeling conditional quantiles rather than averages, quantile regression modeling seeks to link a response variable with a group of regressors. This method works particularly well when there are outliers or skewness in the response. The quantile function for the ATUW distribution is accessible in closed form as

$$x_u(\alpha,\beta) = \frac{2}{\pi}\arctan\left[\left(-\frac{1}{\alpha}\log(1-u)\right)^{\frac{1}{\beta}}\right].$$

This explicit form enables re-parameterization of the distribution in terms ofany quantile $\mu = x_U(\alpha,\beta)$, μ ∈ (0, 1) denote the u^{th} quantile and defining

$$\alpha = -\frac{\log(1-u)}{\left(\tan\left(\frac{\pi\mu}{2}\right)\right)^{\beta}}.$$

We obtain the quantile ATUW distribution with pdf

$$g(y;\beta,\mu,u) = \beta(-\log(1-u))*$$

$$\frac{\left(\tan\left(\frac{\pi y}{2}\right)\right)^{\beta-1}}{\left(\tan\left(\frac{\pi \mu}{2}\right)\right)^{\beta}}\frac{\pi}{2}sec^2\left(\frac{\pi y}{2}\right)*(1-u)^{\left(\frac{\tan\left(\frac{\pi y}{2}\right)}{\tan\left(\frac{\pi \mu}{2}\right)}\right)^{\beta}};$$

$$y \epsilon (0,1).$$

Quantile Regression Model

Suppose $Y_1, Y_2, \cdots, Y_n$ are independent observations such that $Y_i \sim$ QATUW(β,μ_i,u),

Where μ_i depends on the covariates

$x_i=(x_1, x_2, \cdots, x_{ip})$ through a link function

$g(\mu) = {x_i}^T\delta,$

using the logit link,

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right).$$

This model offers a strong substitute for mean regression and is particularly appropriate for skewed or constrained data since it directly links the conditional u^{th}quantile of Y to variables.
To estimate parameters θ = (β, δ), the log-likelihood derived from the QATUW pdf is maximized. Robust quantile regression analysis and flexible modeling of bounded data are made possible by the closed-form quantile function, which makes computation and understanding easier. The ATUW distribution's usefulness is increased by this quantile-based method, which makes it an effective tool for simulating unit interval responses with skewness or heteroscedasticity, which are prevalent in many applicable domains.

III. Results

Simulation Study for Parameter Estimation

A comprehensive simulation study was conducted to evaluate the performance of six parameter estimation methods for the ATUW. These methods MLE, MPSE, LSE, WLSE, ADE, and CVME. Samples were generated from the ATUW with predetermined parameter values (e.g., (α=3, β=5) and (α =5, β=2)) across a range of sample sizes varying from 20 to 1000. For each scenario, 1000 independent replications were performed to ensure the reliability of the results.

The estimators were assessed using key metrics AE, bias, and MSE. These metrics measure the accuracy and precision in estimating the true parameters. The results demonstrate that all estimation methods yield consistent estimates of α and β with both bias and MSE decreasing as sample size increases. Notably, MLE and MPSE showed relatively better performance with smaller bias and variance, especially when sample sizes were moderate to large.

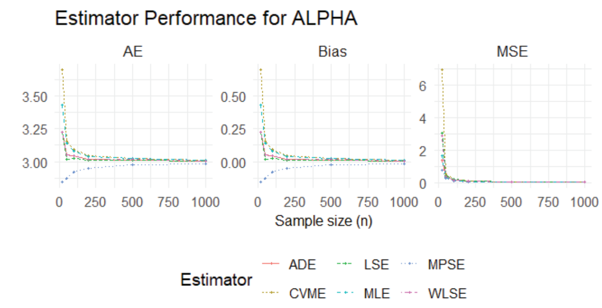

Fig. 1. Estimator of Alpha (α=3)

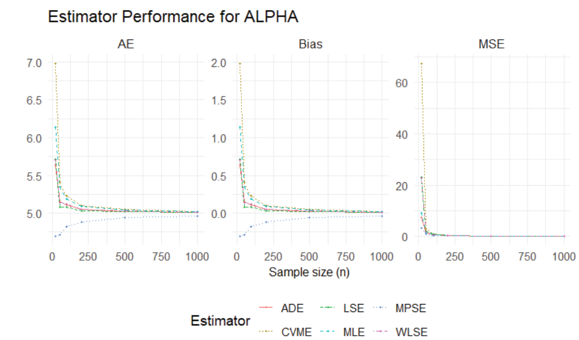

Fig. 2. Estimator of Alpha (α=5)

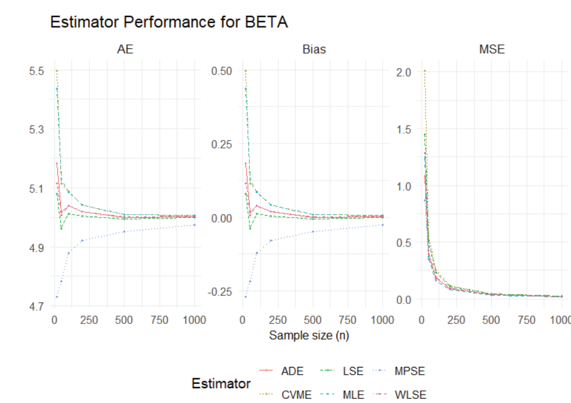

Fig. 3. Estimator of Beta (β=5)

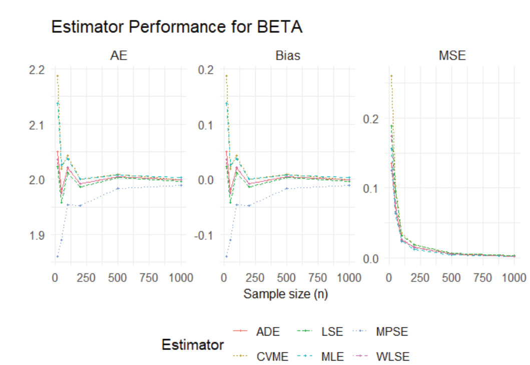

Fig. 4. Estimator of Beta (β=2)

In conclusion, the simulation study confirms the theoretical properties and practical applicability of the proposed estimation methods for the ATUW. While differences among estimators are more evident with small sample sizes, MLE and MPSE are recommended for their robustness and efficiency. These findings establish ATUW as a flexible and estimable model for bounded data, suitable for various applied statistical analyzes.

univariate data application

To evaluate the practical performance of the ATUW, we used a univariate dataset of size n=64, where each observation represents the proportion of children who completed higher levels of education in a district of Bangladesh. These proportions were derived by transforming the original district-level data into unit-scaled values.Summary statistics of the generated data, including minimum, maximum, mean, variance, skewness, and kurtosis, are reported in Table~1.

Subsequently, the ATUW model was fitted to the simulated data using maximum likelihood estimation, alongside the Beta and Kumaraswamy distributions as competitors.

Table 1. Summary Statistics of the Data

Statistic	Minimum	Maximum	Mean	Variance	Skewness	Kurtosis
Value	0.034	0.894	0.501	0.034	-0.178	2.876

Parameter estimates, standard errors, log-likelihoods, and fit statistics (AIC, BIC, Anderson-Darling, Kolmogorov-Smirnov, and Cramér-von Mises) were computed for all models see Table-2.

Table 2. Parameter estimates and goodness-of-fit

Dist	α/a	β/b	SE(α/a)	SE(β/b)	LogLik	AIC	BIC	AD	KS	CvM
ATUW	1.53	2.08	0.28	0.21	35.71	-67.43	-64.63	0.219	0.078	0.036
Beta	1.40	2.18	0.25	0.24	35.29	-66.58	-63.78	0.233	0.085	0.041
Kuma	1.42	2.15	0.30	0.25	35.15	-66.30	-63.50	0.242	0.090	0.044

Goodness-of-fit was further examined through graphical overlays of fitted PDFs and CDFs, Q-Q plots, and P-P plots, all of which demonstrate the superior flexibility and fitting performance of the ATUW distribution for data on the unit interval. Here the PDFs and CDFs, QQ plots, and PP plots are given below:

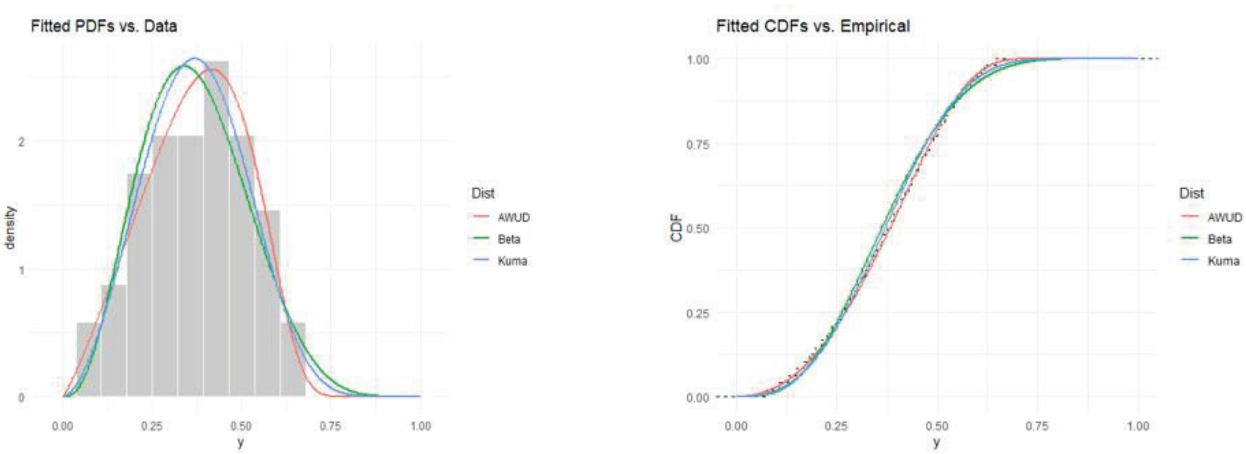

Fig. 6. Fitted PDF and CDF line over data

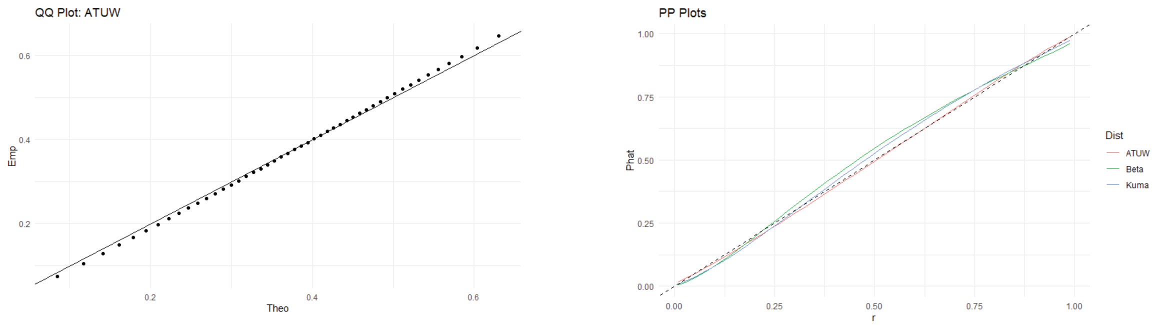

Fig. 7. Q-Q and P-P plot

Data set application with QATUW distribution with comparison other distribution

the parameter estimates and model comparison metrics for the Beta, Kw, LEEG, and QATUW regression models. Among them, the QATUW regression model stands out by presenting the lowest AIC and BIC values along with the highest log-likelihood, indicating that it provides the best overall fit to the data.

In the QATUW model, all parameters are statistically significant at standard significance levels except δ_1 for the reading book to child variable, which does not significantly impact the median response variable. The parameter δ_2 has a positive and statistically significant effect, suggesting that higher maternal education levels are associated with a stronger desire for academic success among adolescents. Similarly, the parameter $\delta_3$significant and positive, indicating that older age at school entry correlates with greater academic ambition. Among the earlier models, the Beta regression model shows moderate fit but includes non-significant parameters (e.g., δ_1). The Kw model improves slightly on fit statistics but also includes non-significant estimates. The LEEG model performs well and has several significant terms, but does not outperform QATUW in AIC/BIC. Thus, the QATUW model not only performs best statistically but also provides the most consistent set of significant predictors, highlighting important socio-educational factors that influence adolescents' motivation for academic achievement. Notably, adolescents who start school later or whose mothers have higher education levels are more likely to aim for top grades, while reading performance alone is not a decisive factor in this particular regression setting.

Table 3. Model Comparison

Model	Parameter	Estimate	SE	p-value	LogLik	AIC	BIC
	β (beta)	7.4359	0.6940	<0.001			
	δ_0 (delta0)	7.7571	1.3654	<0.001			
Beta	δ_1 (delta1)	0.3248	0.3057	0.2882	145.5642	-281.128	-270.334
	δ_2 (delta2)	3.2957	1.1468	0.0041			
	δ_3 (delta3)	0.5859	0.1294	<0.001			
	β (beta)	7.9575	1.4563	<0.001			
	δ_0 (delta0)	7.6115	0.0458	0.3434			
Kw	δ_1 (delta1)	-0.3241	1.3657	0.5342	146.0922	-282.184	-271.390
	δ_2 (delta2)	3.3302	0.0292	<0.001			
	δ_3 (delta3)	0.5720	0.0242	<0.001			
	β (beta)	13.3396	1.3618	<0.001			
	δ_0 (delta0)	-6.9271	0.0754	0.0245			
LEEG	δ_1 (delta1)	-0.4217	0.3881	0.2772	149.1640	-288.328	-277.533
	δ_2 (delta2)	3.2421	0.7792	<0.001			
	δ_3 (delta3)	0.5155	0.1123	0.0258			
	β (beta)	7.4359	0.0440	<0.001			
	δ_0 (delta0)	-7.4706	1.0974	<0.001			
QATUW	δ_1 (delta1)	-0.2995	0.3693	0.4174	149.2989	-288.597	-277.803
	δ_2 (delta2)	3.0382	0.7509	<0.001			
	δ_3 (delta3)	0.5589	0.1029	<0.001			

IV. Conclusion

This study introduces a new unit distribution and quantile regression model tailored for proportion and percentage data, and applies it to educational outcomes in Bangladesh. The comparative results demonstrate that the proposed approach fits the data better than conventional regression models, particularly when analyzing the median of shifted unit response variables.

Among the explanatory factors considered, maternal educational attainment emerges as the most influential determinant of children's higher educational achievement, independent of the age at which schooling began. In contrast, whether someone regularly read books to the child showed no statistically significant effect, a finding that deserves further attention.

Overall, the results highlight the complex interplay of household and early-life factors in shaping educational success. The proposed quantile regression model not only strengthens the statistical modeling of unit-valued responses in education but also provides a versatile tool for future research in other domains where bounded data arise. Importantly, the findings emphasize that policies aiming to expand maternal education and promote timely school enrollment can play a critical role in improving long-term educational attainment across districts.

Acknowledgement

Sincere appreciation is extended to the Bangladesh Bureau of Statistics (BBS) and UNICEF for granting access to the Bangladesh Multiple Indicator Cluster Survey (MICS) 2019 data.

References

1. Koenker, R. 1978. Quantile regression. Econometrica, **46(1)**, 33-50.
2. Ferrari, S. L., Cribari-Neto, F. 2004. Beta regression for modeling rates and proportions. Journal of Applied Statistics, **31(7)**, 799-815.
3. Rangvid, B. S. 2008. School composition effects in Denmark: Quantile regression evidence from PISA 2000. Empirical Economics, **35**, 349-378.
4. Eide, E., D. J., Brewer, R. G. Ehrenberg, 1998. Does it pay to attend an elite private college? Journal of Human Resources, **33(4)**, 704-719.
5. Buchinsky, M. 1988. Changes in the US wage structure 1963–1987: Application of quantile regression. Econometrica, **62(2)**, 405-458.
6. Weibull, W. 1951. A statistical distribution function of wide applicability. Journal of Applied Mechanics, **18**, 293-297.
7. Korkmaz, M. C., et al. 2021. The arc secant hyperbolic normal distribution. Communications in Statistics - Theory and Methods.
8. Kumaraswamy, P. 1980. A generalized probability density function. Journal of Hydrology, **46(1-2)**, 79-88.
9. Govt. of the People's Republic of Bangladesh, BBS and UNICEF, multiple indicator cluster survey 2019: survey findings reported 2019, Dec.2019.Available at: https://www.unicef.org/bangladesh/media/3281/file/Bangladesh%202019%20MICS%20Report_English.pdf