

Article

Not peer-reviewed version

Choosing Between Link Functions for Median Parametric Regression Model Utilizing Median Base Unit Rayleigh Distribution

[Iman Attia](#)*

Posted Date: 16 October 2025

doi: 10.20944/preprints202510.1272.v1

Keywords: parametric quantile regression models; median based unit Rayleigh (MBUR) distribution; logit link function; clog-log function; Nealder mead optimizer; MLE



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Choosing Between Link Functions for Median Parametric Regression Model Utilizing Median Base Unit Rayleigh Distribution

Iman M. Attia

Department of Mathematical Statistics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt; Imanattiathesis1972@gmail.com

Abstract

This study vividly explores the realm of parametric quantile regression through the lens of the innovative Median Based Unit Rayleigh (MBUR) distribution, a cutting-edge one-parameter model that opens new avenues for statistical analysis. The estimation process unfolds through a sleekly re-parameterized maximum likelihood function, illustrated with an engaging real-world dataset that breathes life into the theory. In addition, the author embarks on a thorough investigation into inference and goodness of fit, providing a vivid tapestry of insights that highlights the strength and versatility of the MBUR distribution. The potential of the MBUR distribution to transform the analytical approach is widely recognized and embraced.

Keywords: parametric quantile regression models; median based unit Rayleigh (MBUR) distribution; logit link function; clog-log function; Nealder mead optimizer; MLE

Introduction

Exploring the relationships between covariates and response variables that have highly skewed distributions can be effectively assessed using quantile regression. When the distribution has a well-defined, closed-form quantile function, it can be utilized to reparameterize the probability density function (pdf) and, subsequently, the log-likelihood function. In such cases, parametric quantile regression serves as a strong alternative to least squares regression models, especially when many underlying assumptions may be violated. Some common assumptions include normality and homoscedasticity. Additionally, the presence of outliers can interfere with the estimation of regression coefficients and the inference process. In this context, parametric quantile regression proves to be robust against many of these issues. The literature is abundant with works from various authors discussing quantile regression and its diverse applications across many fields. The median is more robust to outliers and skewness in data compared to the mean; therefore, conditioning on the median rather than the mean is preferable in regression models that exhibit these anomalies. However, any quantile can be utilized. The transmuted Unit Rayleigh distribution, introduced by Korkmaz et al. in 2021, offers a viable alternative to beta and Kumaraswamy regression models, (Korkmaz et al., 2023). Unit-Gompertz was used for modeling inter-record times by (Kumar et al., 2020). (Mazucheli et al., 2023) used its quantile regression to model bounded response data. Mazucheli made tremendous contributions to quantile regression models, particularly in unit distributions such as unit Weibull regression models, (Mazucheli et al., 2020), unit Birnbaum-saunders (Mazucheli et al., 2021), Vesicek quantile and mean regression models for proportional data (Mazucheli et al., 2022). An application of one-parameter unit Lindley (Mazucheli et al., 2019) was introduced by Mazucheli in 2019.

Not only did Mazucheli have these contributions, but others also had immense roles and achievements. (Noufaily & Jones, 2013) has elaborated on the generalized gamma distribution and its parametric quantile regression model since his Ph.D. thesis in 2011 and in a published paper since

2013. Korkmaz clarified and examined numerous unit distributions and their corresponding quantile regression models, (Korkmaz & Chesneau, 2021), (Korkmaz et al., 2023), (Korkmaz et al., 2021). Leiva et al. have provided clear descriptions of the Birnbaum-Saunders distributions since 2016 and continuing thereafter, (Leiva, 2016). Many authors have also provided a detailed explanation of this distribution like (Sánchez, Leiva, Galea, et al., 2021), (Sánchez et al., 2020), (Garcia-Papani et al., 2018). Diving into the fascinating world of quantile regression models uncovers a treasure trove of diverse distributions that can significantly enhance the analytical toolkit. These insights are not only invaluable but also offer powerful new ways to understand the data, (Marchant et al., 2016), (Leão et al., 2018), (Leiva et al., 2023). The authors have developed packages in R that implement quantile regression. In 2020, Mazucheli and Alves created the Vasicekreg package. They followed this with the development of the Ugomquantreg package in 2021. Additionally, Mazucheli and Menezes invented the unitBSQuantReg package in 2020, while Menezes also developed the UWquantreg package that same year. In 2021, Koenker introduced the quantreg package.

In this appendix, the author embarks on an exploration of the innovative Median-based unit Rayleigh (MBUR) distribution discussed previously by Attia (Attia, 2024). This novel approach is intricately designed for conducting quantile regression analysis, providing powerful insights into real-world data applications. The author skillfully illustrates the practical utility of MBUR, offering a fresh perspective that bridges advanced statistical theory with tangible, impactful results in data analysis.

Section One: Proposed Parametric Quantile Regression Model

1.1. Logit Link Function

The response variable follows a MBUR distribution. To establish a causal relationship between this variable and the covariates that may be influencing it, it is crucial to properly specify both the parametric function and the link function. Given that the response variable may be highly skewed and could violate the assumptions of normality and homoscedasticity, using parametric quantile regression may offer a robust solution. However, it is important to explore all other potential techniques to achieve the best estimation results. The MBUR has the following PDF and CDF respectively as expressed in equations (1-2):

$$f(y) = \frac{6}{\alpha^2} \left[1 - y^{\frac{1}{\alpha^2}} \right] y^{\left(\frac{2}{\alpha^2}-1\right)}, 0 < y < 1, \alpha > 0 \quad (1)$$

$$F(y) = 3y^{\frac{2}{\alpha^2}} - 2y^{\frac{3}{\alpha^2}}, 0 < y < 1, \alpha > 0 \quad (2)$$

Quantile regression depends on the quantile function in equation (3):

$$y = F^{-1}(y) = \left\{ -0.5 \left(\cos \left[\frac{\cos^{-1}(1-2u)}{3} \right] - \sqrt{3} \sin \left[\frac{\cos^{-1}(1-2u)}{3} \right] \right) + 0.5 \right\}^{\alpha^2} \quad (3)$$

Re-parameterize the PDF and CDF of MBUR using the quantile function, $y = c^{\alpha^2}$, Where c is $c = -0.5 \left(\cos \left[\frac{\cos^{-1}(1-2u)}{3} \right] - \sqrt{3} \sin \left[\frac{\cos^{-1}(1-2u)}{3} \right] \right) + 0.5$, $u \in (0,1)$. U represents the chosen quantile, if it is the median, so $u=0.5$ so $\alpha^2 = \frac{\ln(y)}{\ln(c)}$. When replacing $u=0.5$ in c , this gives $\ln(0.5) = -0.6931$, $\alpha^2 = \frac{\ln(y)}{\ln(c)} = \frac{\ln(\text{median})}{\ln(c)}$. As y is the median corresponding to $u=0.5$.

Using the logit function of the median, which will be called μ

$$\text{Logit}(\text{median}) = \text{logit}(\mu) = \varphi = X'B = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}; i = 1, \dots, n \ \& \ k = 1, \dots, k$$

where n is the number of cases or observations and k is the number of variables. Logit median is the linear combination of variables.

$$\text{So } \mu = \frac{e^{X'B}}{1+e^{X'B}} = \frac{e^{\varphi}}{1+e^{\varphi}} \text{ and } \alpha^2 = \frac{\ln(y)}{\ln(c)} = \frac{\ln(\text{median})}{\ln(c)} = \frac{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}{\ln(c)} = \frac{\ln\left(\frac{e^{\varphi}}{1+e^{\varphi}}\right)}{\ln(c)}$$

$$\text{The reparameterized PDF is } f(y) = \frac{6\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} \left[1 - y^{\frac{\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}} \right] y^{\frac{2\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} - 1}, 0 < y < 1, \alpha > 0$$

The reparameterized CDF is $F(y) = 3y \frac{2\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} - 2y \frac{3\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}, 0 < y < 1, \alpha > 0$
 The reparameterized quantile is $F^{-1}(y) = \left\{ -0.5 \left(\cos \left[\frac{\cos^{-1}(1-2u)}{3} \right] - \sqrt{3} \sin \left[\frac{\cos^{-1}(1-2u)}{3} \right] \right) + \frac{\ln\left(\frac{e^\varphi}{1+e^\varphi}\right)}{\ln(c)} \right\}$

1.1.1. Estimation Using Log-Likelihood Function

Re-parameterize the PDF and subsequently the log-likelihood by replacing this α^2 in the log-likelihood function to estimate the parameter using MLE as shown in equations (4-5):

$$l(\alpha) = n\ln(6) - n\ln(\alpha^2) + \sum_{i=1}^n \ln \left[1 - y_i^{\alpha^2} \right] + \sum_{i=1}^n \left(\frac{2}{\alpha^2} - 1 \right) \ln(y_i) \quad (4)$$

$$l(\alpha) = n\ln(6) - n\ln \left(\frac{\ln\left(\frac{e^\varphi}{1+e^\varphi}\right)}{\ln(c)} \right) + \sum_{i=1}^n \ln \left[1 - y_i^{\frac{\ln(c)}{\ln\left(\frac{e^\varphi}{1+e^\varphi}\right)}} \right] + \sum_{i=1}^n \left(\frac{2 \ln(c)}{\ln\left(\frac{e^\varphi}{1+e^\varphi}\right)} - 1 \right) \ln(y_i) \quad (5)$$

$$l(\beta) = n\ln(6) - n\ln \left(\frac{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}{\ln(c)} \right) + \sum_{i=1}^n \ln \left[1 - y_i^{\frac{\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}} \right] + \sum_{i=1}^n \left(\frac{2 \ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} - 1 \right) \ln(y_i)$$

Next, apply the chain rule to differentiate the log-likelihood for estimating the regression coefficients. The log-likelihood is divided into 3 terms.

$$f_1(\beta) = \ln \left(\frac{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}{\ln(c)} \right), f_2(\beta) = \ln \left[1 - y_i^{\frac{\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}} \right], f_3(\beta) = \left(\frac{2 \ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} - 1 \right) \ln(y_i)$$

Each one of the above terms will be differentiated with respect to each of the regression coefficients using the chain rule.

$$\frac{\partial l(\beta)}{\partial \beta_k} = \frac{\partial l(\beta)}{\partial \mu} \frac{\partial \mu}{\partial \varphi} \frac{\partial \varphi}{\partial \beta}$$

The linear regression equation with any number of variables can be expressed as

$$\text{logit}(\mu) = \varphi = X'B = \beta_0 + \beta_k x_{ik} \text{ and hence } \mu = \frac{e^\varphi}{1+e^\varphi} = \frac{e^{\beta_0 + \beta_k x_{ik}}}{1+e^{\beta_0 + \beta_k x_{ik}}}$$

Start with: differentiate the first term w.r.t. to both regression coefficients in equations (6)

$$\frac{\partial f_1(\beta)}{\partial \beta_0} = \frac{\frac{\partial}{\partial \beta} \left(\frac{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}{\ln(c)} \right)}{\frac{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}{\ln(c)}} = \frac{1}{(1+e^\varphi)\ln(\mu)} \quad \& \quad \frac{\partial f_1(\beta)}{\partial \beta_k} = \frac{x_{ik}}{(1+e^\varphi)\ln(\mu)} \quad (6)$$

Next, differentiate the second term w.r.t. to both regression coefficients in equations (7)

$$\frac{\partial}{\partial \beta_0} f_2(\beta) = \frac{\partial}{\partial \beta_0} \ln \left[1 - y_i^{\frac{\ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)}} \right] = \frac{\partial}{\partial \beta_0} \ln \left[1 - y_i^{\frac{\ln(c)}{\ln(\mu)}} \right] \quad (7)$$

$$\frac{\partial f_2(\beta)}{\partial \beta_0} = \frac{-y_i^{\frac{\ln(c)}{\ln(\mu)}} \left[\ln(y) \right] \left[\frac{-\ln(c)}{(1+e^\varphi) [\ln(\mu)]^2} \right]}{\left[1 - y_i^{\frac{\ln(c)}{\ln(\mu)}} \right]} \quad \& \quad \frac{\partial f_2(\beta)}{\partial \beta_k} = \frac{-y_i^{\frac{\ln(c)}{\ln(\mu)}} \left[\ln(y) \right] \left[\frac{-x_{ik} \ln(c)}{(1+e^\varphi) [\ln(\mu)]^2} \right]}{\left[1 - y_i^{\frac{\ln(c)}{\ln(\mu)}} \right]}$$

Lastly, differentiate the third term w.r.t. to both regression coefficients

$$\frac{\partial f_3(\beta)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \left(\frac{2 \ln(c)}{\ln\left(\frac{e^{X'B}}{1+e^{X'B}}\right)} - 1 \right) \ln y_i = \frac{-2 \ln(c) \ln y_i}{(1+e^\varphi) [\ln(\mu)]^2} \quad \& \quad \frac{\partial f_3(\beta)}{\partial \beta_k} = \frac{-2 x_{ik} \ln(c) \ln y_i}{(1+e^\varphi) [\ln(\mu)]^2}$$

These equations can be solved using numerical algorithm and an optimizer like the quasi-newton:

$$\frac{\partial l(\beta)}{\partial \beta_0} = \frac{-1}{(1+e^\varphi)\ln(\mu)} + \frac{-y_i \frac{\ln(c)}{\ln(\mu)} [\ln(y_i)] \left[\frac{-\ln(c)}{(1+e^\varphi) [\ln(\mu)]^2} \right]}{\left[1 - y_i \frac{\ln(c)}{\ln(\mu)} \right]} + \frac{-2 \ln(c) \ln(y_i)}{(1+e^\varphi) [\ln(\mu)]^2}$$

$$\frac{\partial l(\beta)}{\partial \beta_k} = \frac{-x_{ik}}{(1+e^\varphi)\ln(\mu)} + \frac{-y_i \frac{\ln(c)}{\ln(\mu)} [\ln(y_i)] \left[\frac{-x_{ik} \ln(c)}{(1+e^\varphi) [\ln(\mu)]^2} \right]}{\left[1 - y_i \frac{\ln(c)}{\ln(\mu)} \right]} + \frac{-2 x_{ik} \ln(c) \ln(y_i)}{(1+e^\varphi) [\ln(\mu)]^2}$$

Where $i=1,2,\dots,n$ and $k=1,2,\dots,p$ the n are the number of observations while the k are the number of parameters or predictors.

1.2. Log-Log Median Link Function

A link function other than the logit can be used. The author used the log-log function $\log\{-\log(\text{median})\} = X'B = \varphi = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$, $i = 1, 2, \dots, n$ & $k = 1, 2, \dots, k$

$$\text{median} = e^{-e^{X'\beta}} = \exp(-\exp(X'\beta)) \quad \& \quad \alpha^2 = \frac{\ln(\text{median})}{\ln(c)} = \alpha^2 = \frac{\ln(e^{-e^{X'\beta}})}{\ln(c)} = \frac{-e^{X'\beta}}{\ln(c)}$$

Now using this parameter to re-parameterize the log-likelihood and the following functions

$$\text{The reparameterized PDF is } f(y) = \frac{6}{\frac{-e^{X'\beta}}{\ln(c)}} \left[1 - y \frac{\frac{1}{\frac{-e^{X'\beta}}{\ln(c)}}}{\frac{-e^{X'\beta}}{\ln(c)}} \right] y^{\frac{2}{\frac{-e^{X'\beta}}{\ln(c)}} - 1}, \quad 0 < y < 1, \alpha > 0$$

$$\text{The reparameterized CDF is } F(y) = 3y^{\frac{2}{\frac{-e^{X'\beta}}{\ln(c)}}} - 2y^{\frac{3}{\frac{-e^{X'\beta}}{\ln(c)}}}, \quad 0 < y < 1, \alpha > 0$$

$$\text{The reparameterized quantile is: } F^{-1}(y) = \left\{ -0.5 \left(\cos \left[\frac{\cos^{-1}(1-2u)}{3} \right] - \sqrt{3} \sin \left[\frac{\cos^{-1}(1-2u)}{3} \right] \right) + \frac{-e^{X'\beta}}{\ln(c)} \right\}$$

And for this link function the derivatives of the reparameterized log-likelihood function are:

$$\frac{\partial l(\beta)}{\partial \beta_0} = -1 + \frac{-y_i \frac{\ln(c)}{e^{-\varphi}} [\ln(y_i)] [\ln(c)] [e^{-\varphi}] (1)}{\left[1 - y_i \frac{\ln(c)}{e^{-\varphi}} \right]} + 2 \ln(c) \ln(y_i) [e^{-\varphi}] (1)$$

$$\frac{\partial l(\beta)}{\partial \beta_k} = -x_{ik} + \frac{-y_i \frac{\ln(c)}{e^{-\varphi}} [\ln(y_i)] [\ln(c)] [e^{-\varphi}] (x_{ik})}{\left[1 - y_i \frac{\ln(c)}{e^{-\varphi}} \right]} + 2 \ln(c) \ln(y_i) [e^{-\varphi}] x_{ik}$$

1.3. Complementary Log-Log Link Function

$\log\{-\log(1 - \text{median})\} = X'B = \varphi = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$; $i = 1, \dots, n$ & $k = 1, \dots, k$

$$\text{median} = 1 - e^{-e^{X'\beta}} = 1 - \exp(-\exp(X'\beta)) \quad \& \quad \alpha^2 = \frac{\ln(\text{median})}{\ln(c)} = \frac{\ln(1 - e^{-e^\varphi})}{\ln(c)}$$

And for this link function the derivatives of the reparameterized log-likelihood function are:

$$\frac{\partial l(\beta)}{\partial \beta_0} = \frac{-e^{-e^\varphi} e^\varphi}{\ln \mu} + \frac{y_i \frac{\ln(c)}{\ln \mu} [\ln(y_i)] [\ln(c)] [e^{-e^\varphi} e^\varphi]}{\left[1 - y_i \frac{\ln(c)}{\ln \mu} \right] [\mu] [\ln \mu]} - \frac{2 \ln(c) \ln(y_i) [e^{-e^\varphi} e^\varphi]}{\mu [\ln \mu]}$$

$$\frac{\partial l(\beta)}{\partial \beta_k} = \frac{-x_{ik} e^{-e^\varphi} e^\varphi}{\ln \mu} + \frac{y_i \frac{\ln(c)}{\ln \mu} [\ln(y)] [\ln(c)] [e^{-e^\varphi} e^\varphi](x_{ik})}{\left[1 - y_i \frac{\ln(c)}{\ln \mu}\right] [\mu] [\ln \mu]} - \frac{2 \ln(c) \ln(y_i) [e^{-e^\varphi} e^\varphi](x_{ik})}{\mu [\ln \mu]}$$

Section Two

2.1. Goodness of Fit Measures

Model adequacy can be evaluated by analyzing the residuals. In quantile regression, two well-known types of residuals are examined for their distribution.

First type: Randomized quantile (RQ) residuals: $r_i^{RQ} = \Phi^{-1}(F(y_i, \hat{\beta}))$

Φ is the standard normal CDF. F is the re-parameterized MBUR CDF given in equation (6). y_i are the observations and $\hat{\beta}$ are the estimated regression coefficients. These residuals are approximately distributed as standard normal. When the model is correctly specified, these residuals approximately follow the standard normal distribution.

Second type: Cox-Snell (CS) residuals: $r_i^{CS} = -\log(1 - F(y_i, \hat{\beta}))$

Cox-Snell residuals are approximately distributed as a standard exponential distribution with a scale parameter of one. The negative logarithm mentioned above represents the cumulative distribution function (CDF) of the standard exponential distribution with this scale parameter. When the model is correctly specified, these residuals will approximately follow a standard exponential distribution.

2.2. Model Selection Criteria

Models are selected based on the lowest values of AIC, BIC, and Corrected AIC among the available options. (Sánchez, Leiva, Saulo, et al., 2021) used the R_M^2 , which is correspondent to the usual R^2 in mean regression. So it can be used and it is defined as

$$R_M^2 = 1 - \exp\left(\frac{2}{n}(l(\hat{\theta}) - l(\hat{\theta}))\right)$$

where $l(\hat{\theta})$ and $l(\hat{\theta})$ are the maximum log-likelihood for the model without covariates (null model) and the model with all covariates (Full model).

$$AIC = -2l(\hat{\theta}) + 2p, CAIC = AIC + \frac{2p(p+1)}{n-p-1}, BIC = -2l(\hat{\theta}) + p \log(n)$$

where p is the number of parameters and n is the sample size.

Section Three: Real Data Analysis, Results and Discussion

3.1. Descriptive Analysis

Table 1 shows the dataset obtained from the OECD which stands for Organization for Economic Co-operation and Development. MATLAB was used for analysis. Nealder Mead optimizer was used to estimate the parameters. Finite central difference method was used to calculate the variance.

The data is available at <https://stats.oecd.org/index.aspx?DataSetCode=BLI>

The water quality indicator is expressed as the percentage of population satisfied with the water quality. The air pollution is expressed as the concentration of PM2.5 in the air and is measured in micrograms per cubic meter. Higher concentration indicates worse air quality.

Water quality data values are: 0.92, 0.92, 0.79, 0.9, 0.62, 0.82, 0.87, 0.89, 0.93, 0.86, 0.97, 0.78, 0.91, 0.67, 0.81, 0.97, 0.8, 0.77, 0.77, 0.87, 0.82, 0.83, 0.83, 0.85, 0.75, 0.91, 0.85, 0.98, 0.82, 0.89, 0.81, 0.93, 0.76, 0.97, 0.96, 0.62, 0.82, 0.88, 0.7, 0.62, 0.72. the descriptive statistics of this dataset are; the mean is 0.8332, the standard deviation is 0.0972, the skewness is -0.6059, the kurtosis is 2.9144, the minimum value is 0.62, the maximum value is 0.98, the 25th quantile is 0.7775, the median is 0.83, and the 75th quantile is 0.91. The data are negatively skewed (left skewness) and platykurtic.

Air pollution data values are: 6.7, 12.2, 12.8, 7.1, 23.4, 22.6, 17.5, 17, 10, 5.9, 5.5, 11.4, 12, 14.5, 16.7, 6.4, 7.8, 19.7, 15.9, 13.7, 27.3, 12.7, 10.5, 10, 20.3, 12.2, 6, 6.7, 22.8, 8.3, 18.5, 17, 10, 5.8, 10.1, 27.1, 10.1, 7.7, 11.7, 11.8, 28.5. The descriptive statistics of the air pollution data are as follows; the mean is 13.5098, the standard deviation is 6.3942, the skewness is 0.8055, the kurtosis is 2.8203, minimum value is 5.5, the maximum value is 28.5, the 25th quantile is 8.175, the median is 12, and the 75th quantile is 17.125. The data are positively skewed (right skewness) and leptokurtic.

3.2. The Model: Regressing Water Quality on Air Pollution

The predictors are transformed by dividing the predictor by 100 and then taking the log of the result. The Kendall tau between the response variable and the predictor is -0.4144 and it is statistically significant with p value 1.783e-4. The response variable is distributed as Median Based Unit Rayleigh. The author used 3 models, in each model a different link function is utilized and the model is named after the name of the link function. The link functions are: the logit, the log-log, and the clog-log. The parameter is reparametrized and hence the PDF using different link functions. Each model results into estimated median value as well as estimated alpha parameter value for each observation. For each model there is a set of figures illustrating the fitted median vs. the predictors, the predictor against the residuals, the QQ plot for the randomized quantile residuals and the Cox-Snell residuals. Table 1 shows the comparison between the different models according to the link function. For each model, the estimated intercept and slope are recorded. The Wald statistics and the associated p value, the log-likelihood, AIC, CAIC, BIC, and HQIC are recorded. The Likelihood Ratio Test (LRT) and the associated p value are also reported. The R-squared is also shown. The Kendall tau with its p value reflecting the relation between the transformed predictors and the residuals are stated. The standard error can be determined by taking the square root of the diagonal of the variance-covariance matrix which is prescribed in the table.

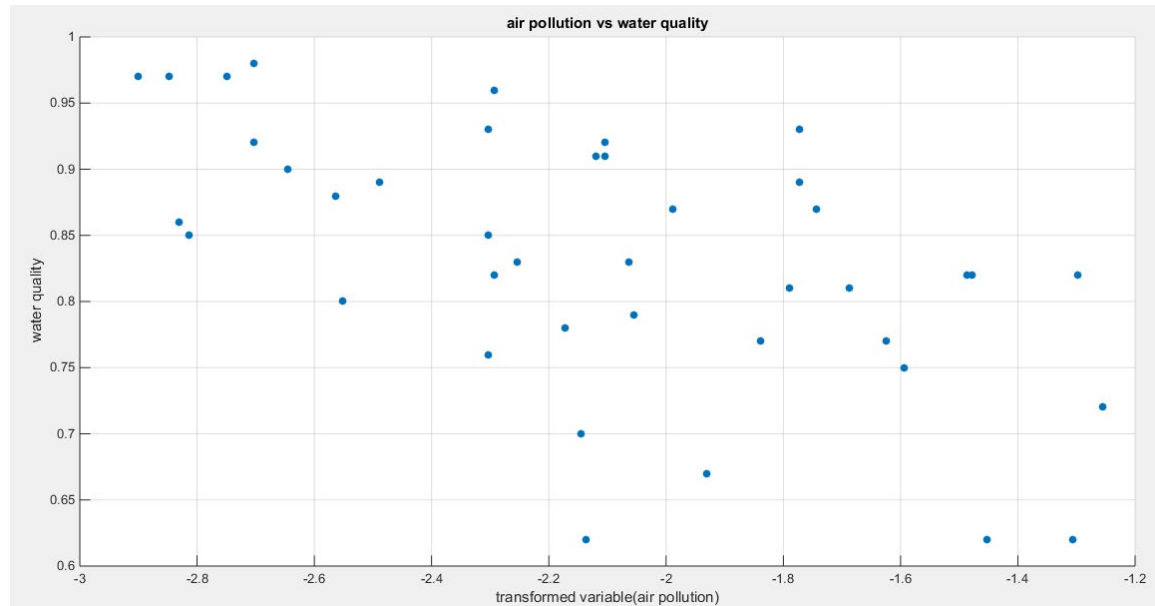


Figure 1. Shows the scatter plot of the water quality against the transformed predictor (air pollution).

Table 1. Regressing water quality on air pollution, comparison between the 3 models.

	Logit link function	Log-log complementary	Log-log median
b0	-0.1521	-0.2213	-0.0937
b1	-0.9411	-0.4233	0.8630
LL	46.2983	46.5111	46.2508
Wald stat. of b0	-0.2498 (p > 0.025)	-0.7878 (p > 0.025)	-0.1704 (p > 0.025)

Wald stat. of b1	-3.3696 (p < 0.025)	-3.4562 (p < 0.025)	3.3798 (p < 0.025)			
AIC	-88.5967	-89.0223	-88.5016			
CAIC	-88.2809	-88.7065	-88.1858			
BIC	-85.1695	-85.5951	-85.0745			
HQIC	-87.3487	-87.7743	-87.2537			
LRT	11.6015(p-val.=0.0007)	12.0271(p-val=0.00052)	11.5064(0.0007)			
R-squared	0.2465	0.2442	0.2447			
P-value for randomized quantile residuals	0.7791	0.8083	0.778			
p-value for Cox-snell residuals	0.7791	0.8083	0.778			
Variance-covariance matrix	0.3706	0.1666	0.0789	0.0339	0.3024	0.1375
	0.1666	0.0780	0.0339	0.015	0.1375	0.0652
QR vs. predictor(t,p)	0.0159,0.8927	0.0159,0.8927	0.0159,0.8927			
CS vs. predictor(t,p)	0.0159,0.8927	0.0159,0.8927	0.0159,0.8927			

Figures for the Logit Model

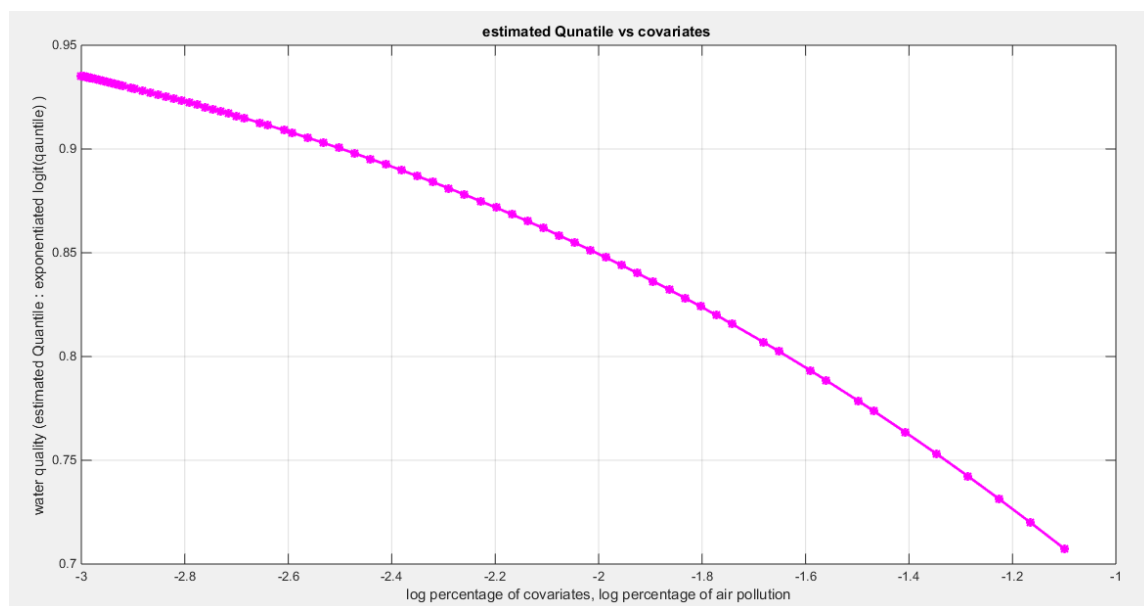


Figure 2. Shows the decreasing estimated curve for the median against the transformed predictor.

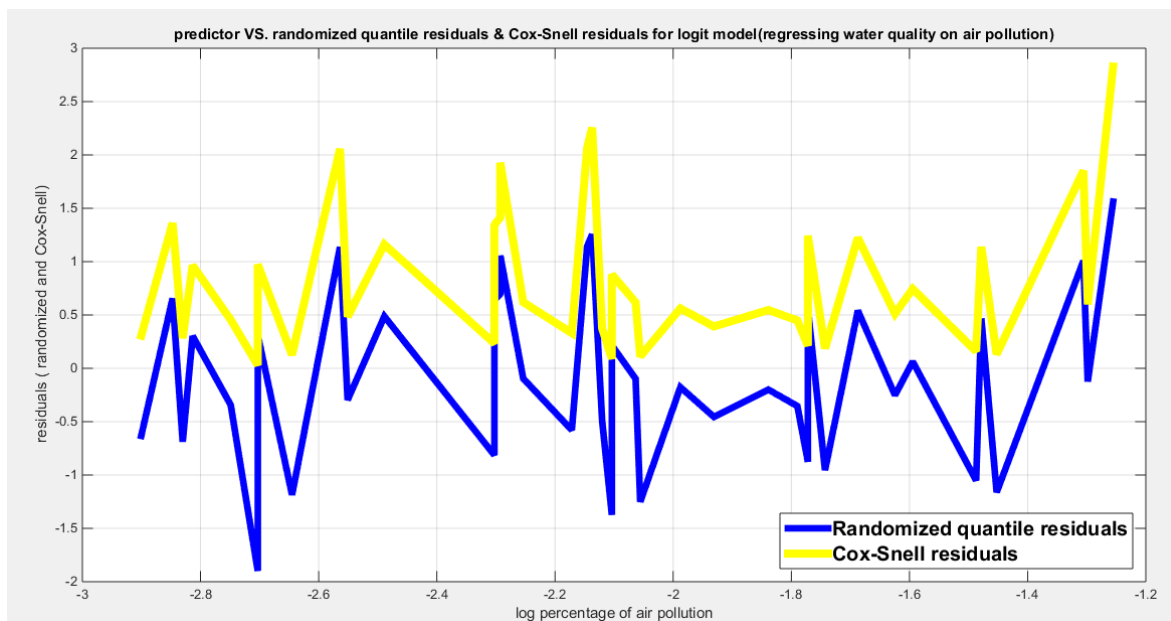


Figure 3. Shows the predictors vs. the residuals of both types. Both curves show no specific trend and they are randomly related.

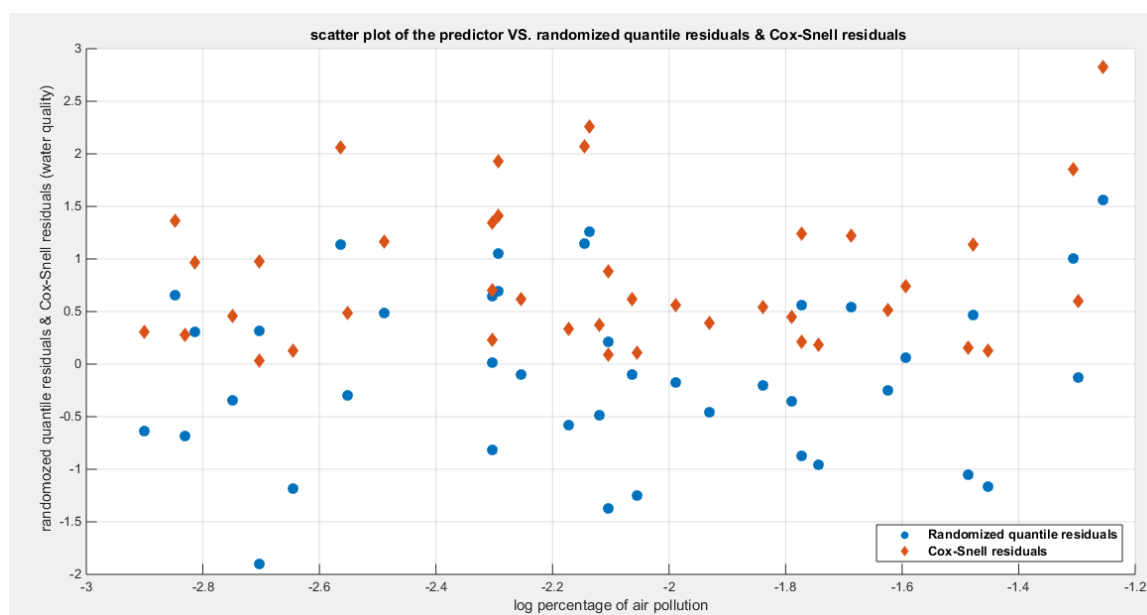


Figure 4. Shows the residuals against the predictors. No specific trend is there and they are randomly scattered. This behavior is clear and obvious for both types of residuals.

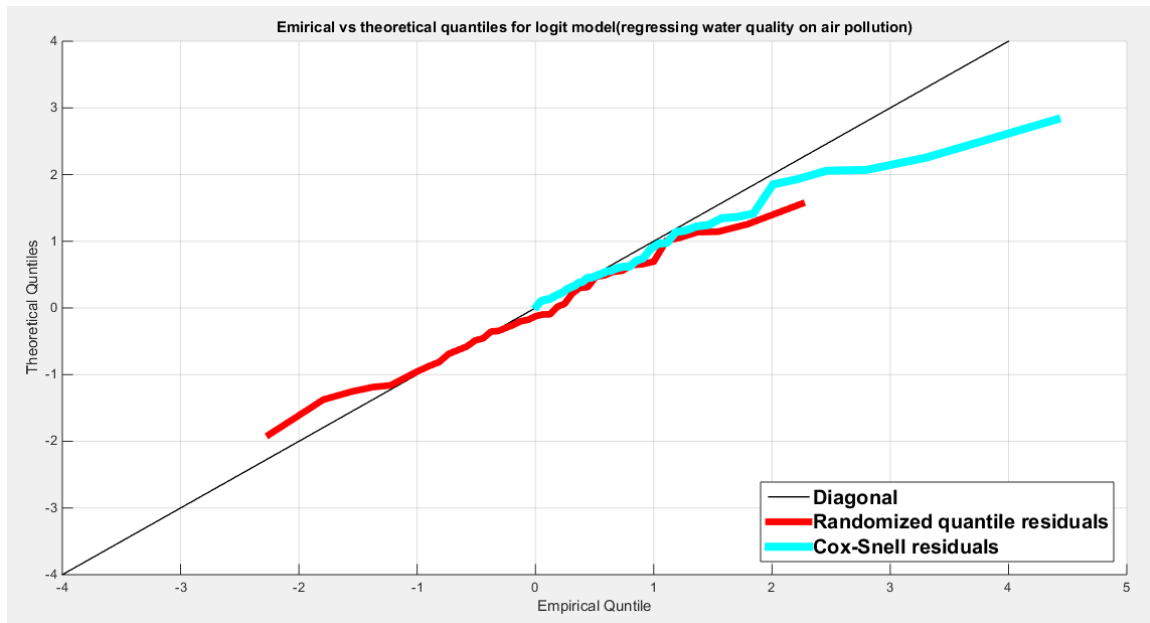


Figure 5. Shows the QQ plot of the empirical quantiles for both residuals . The randomized quantile residuals are more or less perfectly aligned with the diagonal than the Cox Snell residual. Cox-Snell residuals are nearly aligned with the diagonal at the lower end than at the center and the upper tail.

Figures for the Log-log complementary Model

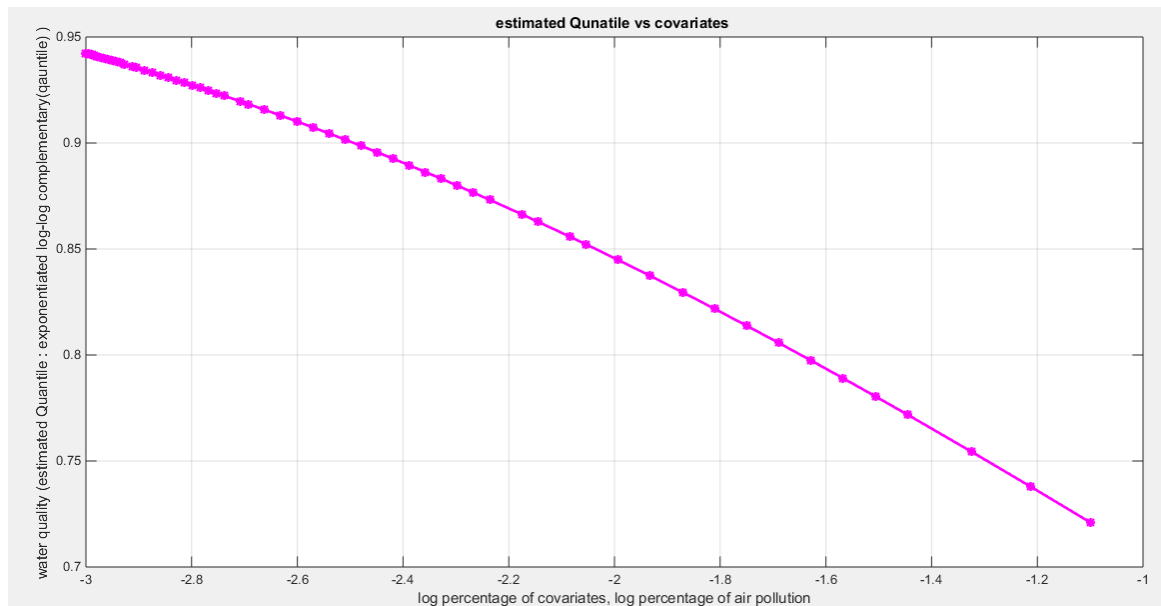


Figure 6. Shows the decreasing estimated curve for the median against the transformed predictor.

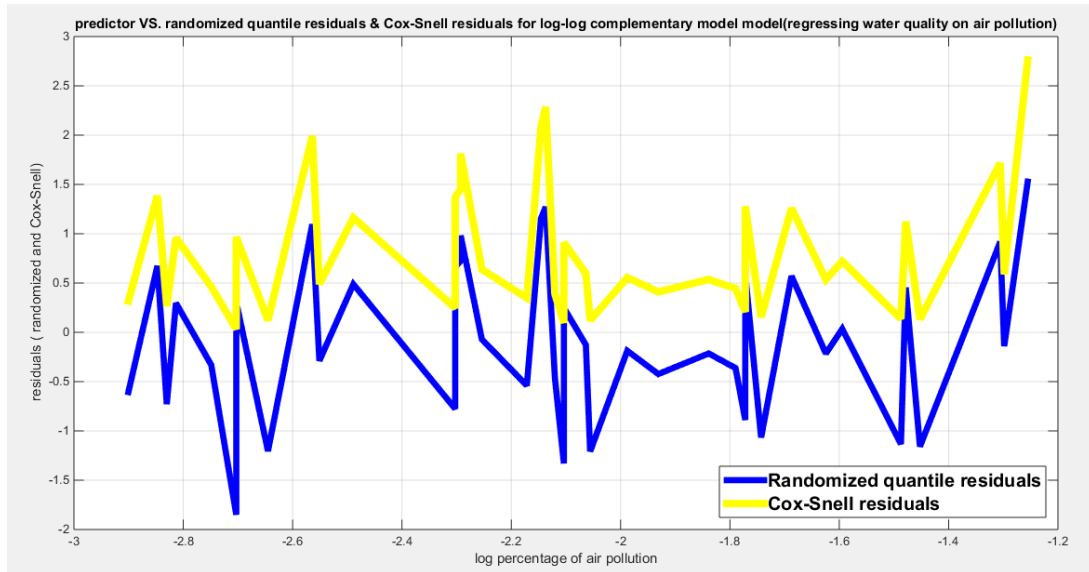


Figure 7. Shows the predictors vs. the residuals of both types. Both curves show no specific trend and they are randomly related.

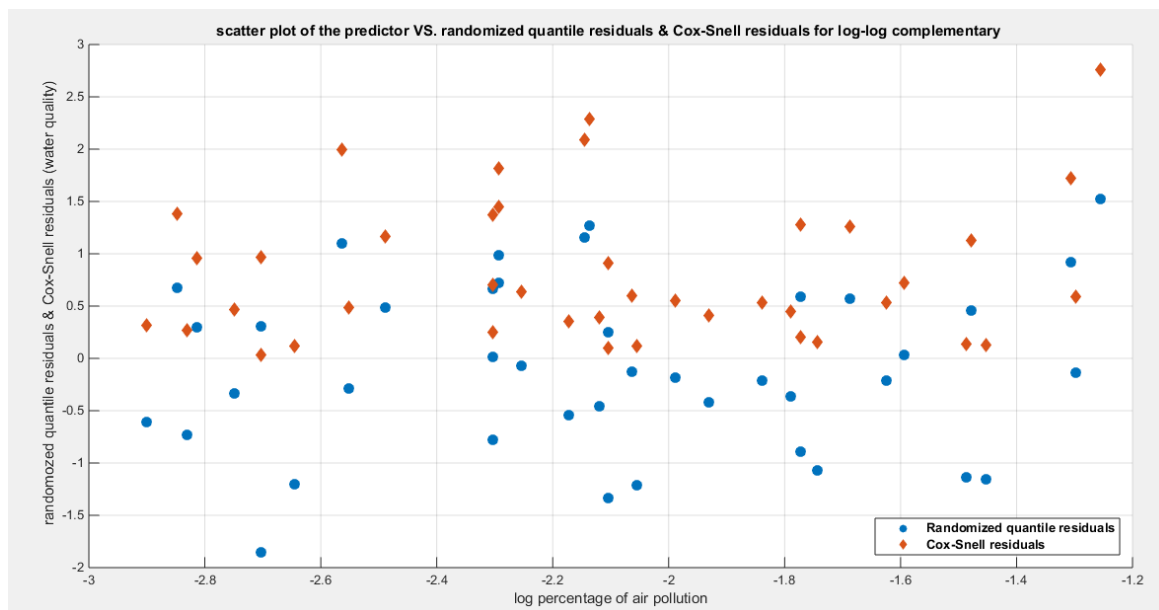


Figure 8. Shows the residuals against the predictors. No specific trend is there and they are randomly scattered. This behavior is clear and obvious for both types of residuals.

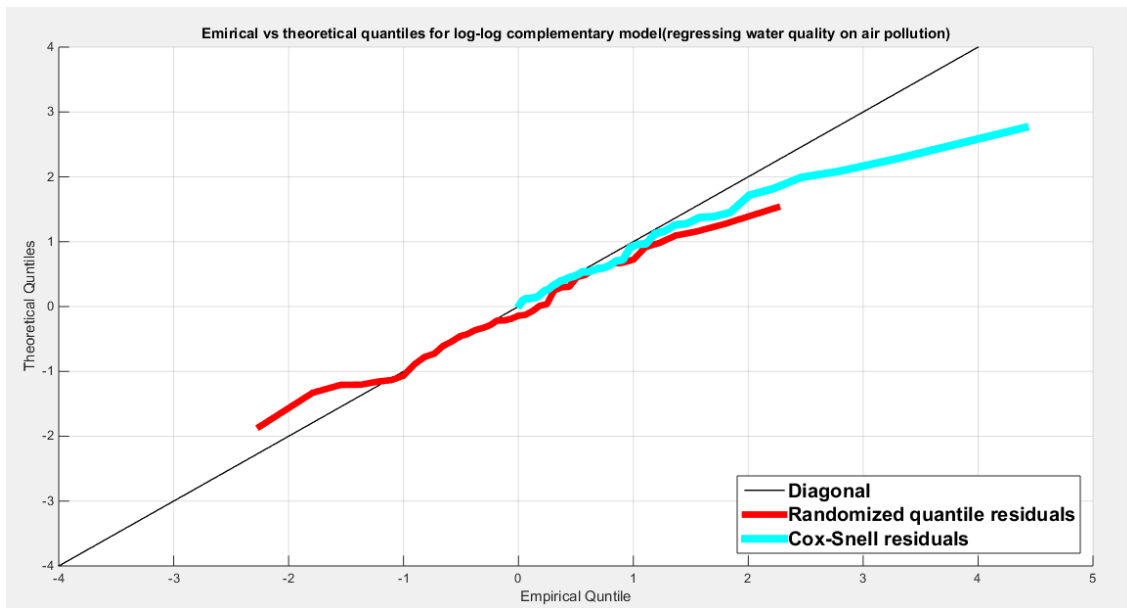


Figure 9. Shows the QQ plot of the empirical quantiles for both residuals . The randomized quantile residuals are more or less perfectly aligned to the diagonal than the Cox Snell residual. Cox-Snell residuals are nearly aligned to diagonal at the lower end than at the center and upper tail.

Figures for the Log-log median Model

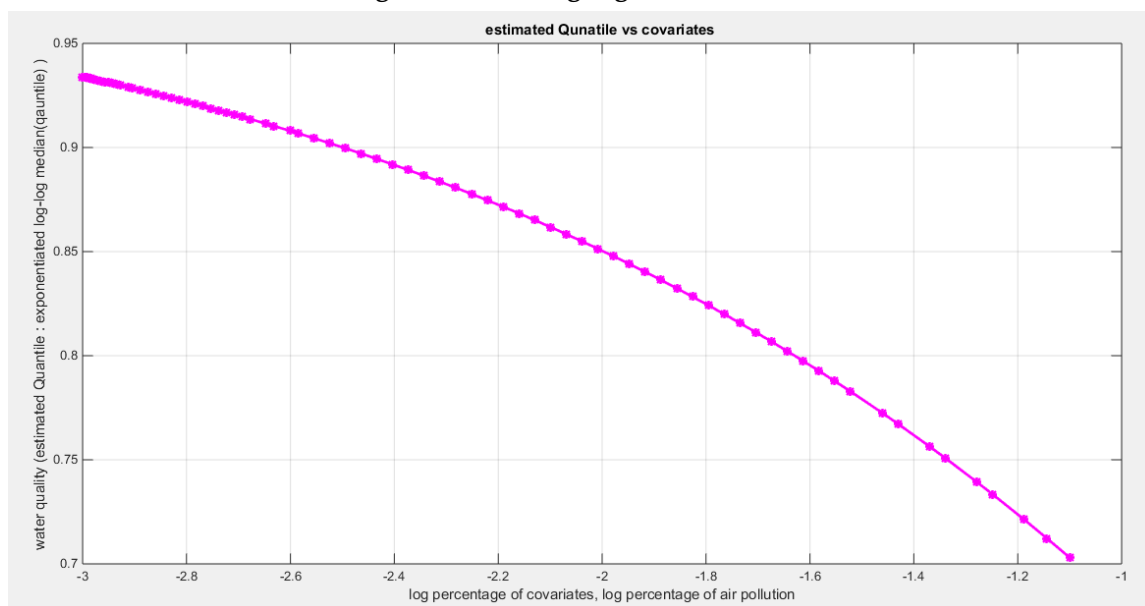


Figure 10. Shows the decreasing estimated curve for the median against the transformed predictor.

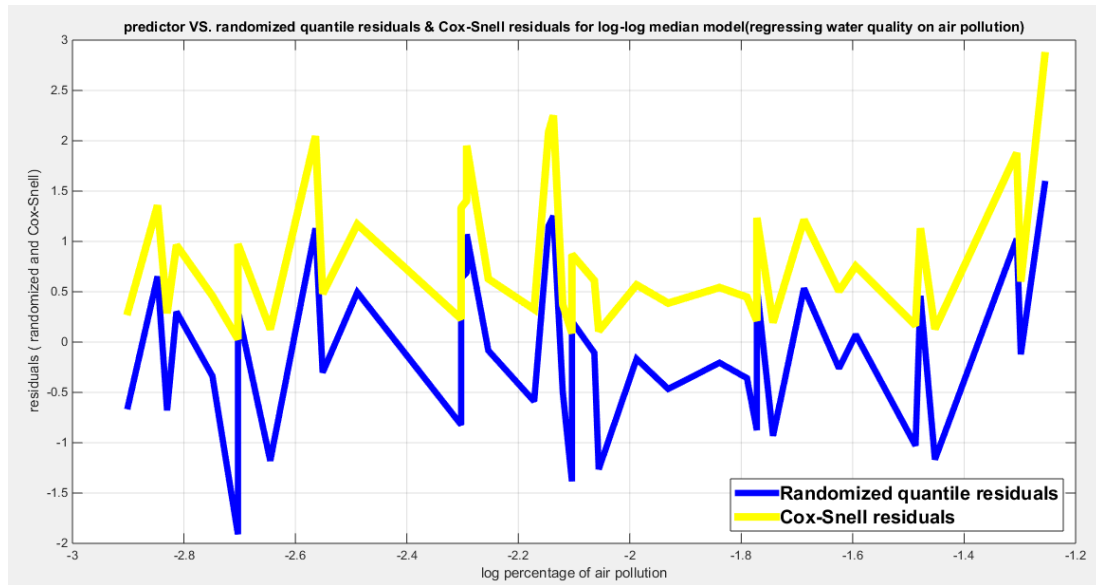


Figure 11. Shows the predictors vs. the residuals of both types. Both curves show no specific trend and they are randomly related.

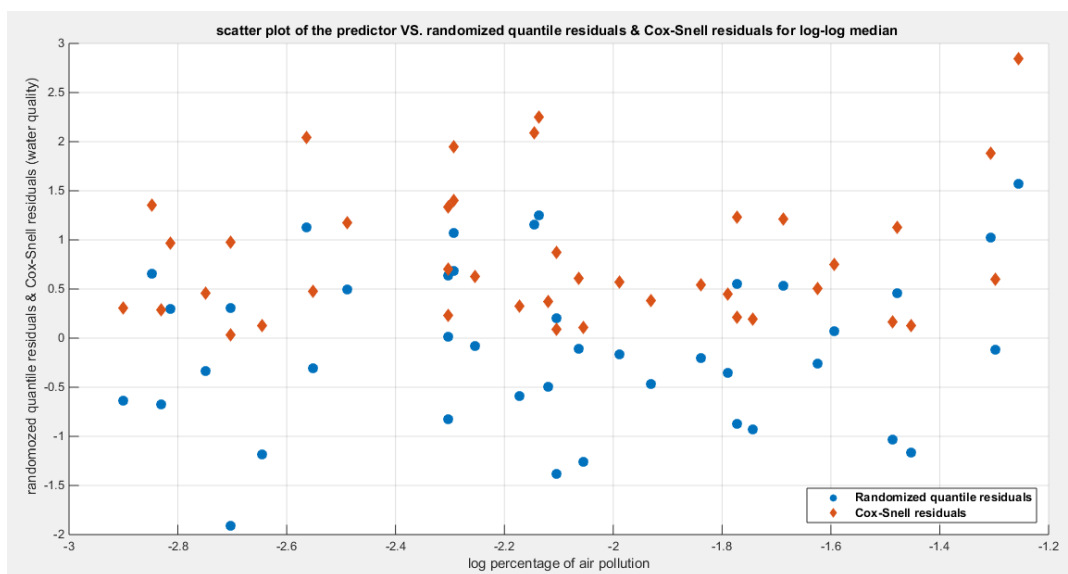


Figure 12. Shows the residuals against the predictors. No specific trend is there and they are randomly scattered. This behavior is clear and obvious for both types of residuals.

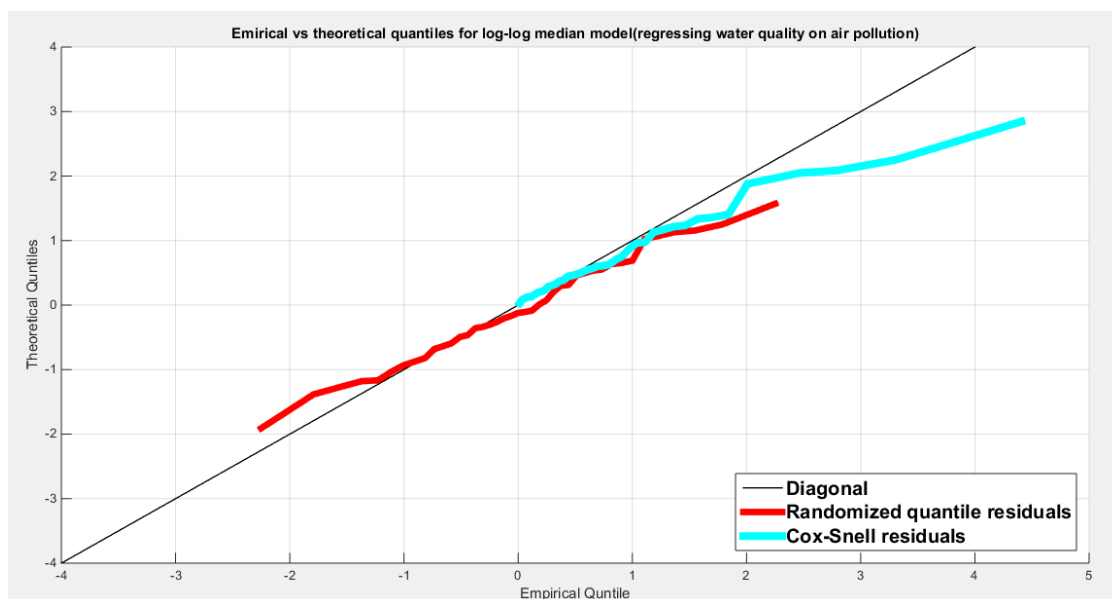


Figure 13. Shows the QQ plot of the empirical quantiles for both residuals. The randomized quantile residuals are more or less perfectly aligned to the diagonal than the Cox Snell residual. Cox-Snell residuals are nearly aligned to diagonal at the lower end than at the center and upper tail.

From the above table and figures, the three models show minimal difference in the statistical indices, they all fit the data and even if the link functions are monotonically increasing or decreasing, the estimated curve is a decreasing function in the predictor. That is to mean, the increased percentage of air pollution decreases the quality of water and this account for about 24 %. Other predictors should be investigated and added to the OECD platform. The coefficient has different sign in the log-log model; it has a positive sign while it has a negative sign in the other model. The QQ plot of the empirical quantiles against the theoretical quantiles shows almost near perfect alignment of the randomized quantile residuals with the diagonal all through its course in comparison of the alignment of the Cox-Snell residuals with the diagonal, which shows near alignment with the diagonal at the lower tail only, as it is not perfectly aligned at the center and the upper tail like the QR. the Kendall tau denotes that the relation between the residuals and the predictors are not statistically significant as the tau is 0.0159 and the associated p value is 0.8927. The plot depicting this relation shows no specific trend and the residuals are randomly scattered all over the different values of the transformed predictors. The figures for the 3 models have similar patterns.

Section Four: Conclusion

Parametric quantile regression emerges as a powerful and resilient approach, especially in scenarios where the foundational assumptions of normality and homoscedasticity falter. It shines when outliers disrupt the integrity of the dataset, offering a robust alternative to traditional methods. In this appendix, the author delves deeply into this compelling subject, illuminating the intricacies of the methodology through the lens of real-life data analysis, thereby illustrating its practical applications and benefits.

Future Work

More rigorous procedures are needed to tackle the issues of non-normality, heteroscedasticity, and presence of outliers. These subjects and other anomilies are challenges that may face the analysis of model regressing responses on predictors exhibiting nonlinear relationship that may be non-monotonically increasing or decreasing.

Authors' Contribution: AI (Attia Iman) carried the conceptualization by formulating the goals, aims of the research article, formal analysis by applying the statistical, mathematical and computational techniques to synthesize and analyze the hypothetical data, carried the methodology by creating the model, software programming and implementation, supervision, writing, drafting, editing, preparation, and creation of the presenting work.

Funding: No funding resource. No funding roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript are declared

Ethics Approval and Consent to Participate: Not applicable.

Consent for Publication: Not applicable

Availability of Data and Material: Not applicable. Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgement: Not applicable

Competing Interests: The author declares no competing interests of any type.

References

1. Attia, I. (2024). *A Novel Unit Distribution Named as Median Based Unit Rayleigh (MBUR): Properties and Estimations*. <https://doi.org/10.20944/preprints202410.0448.v1>
2. Garcia-Papani, F., Leiva, V., Uribe-Opazo, M. A., & Aykroyd, R. G. (2018). Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. *Chemometrics and Intelligent Laboratory Systems*, 177, 114–128. <https://doi.org/10.1016/j.chemolab.2018.03.012>
3. Korkmaz, M. Ç., & Chesneau, C. (2021). On the unit Burr-XII distribution with the quantile regression modeling and applications. *Computational and Applied Mathematics*, 40(1), 29. <https://doi.org/10.1007/s40314-021-01418-5>
4. Korkmaz, M. Ç., Chesneau, C., & Korkmaz, Z. S. (2021). On the Arcsecant Hyperbolic Normal Distribution. Properties, Quantile Regression Modeling and Applications. *Symmetry*, 13(1), 117. <https://doi.org/10.3390/sym13010117>
5. Korkmaz, M. Ç., Chesneau, C., & Korkmaz, Z. S. (2023). A new alternative quantile regression model for the bounded response with educational measurements applications of OECD countries. *Journal of Applied Statistics*, 50(1), 131–154. <https://doi.org/10.1080/02664763.2021.1981834>
6. Kumar, D., Dey, S., Ormoz, E., & MirMostafae, S. M. T. K. (2020). Inference for the unit-Gompertz model based on record values and inter-record times with an application. *Rendiconti Del Circolo Matematico Di Palermo Series 2*, 69(3), 1295–1319. <https://doi.org/10.1007/s12215-019-00471-8>
7. Leão, J., Leiva, V., Saulo, H., & Tomazella, V. (2018). Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. *Statistics in Medicine*, 37(29), 4421–4440. <https://doi.org/10.1002/sim.7929>
8. Leiva, V. (Ed.). (2016). *The Birnbaum-Saunders distribution*. Elsevier.
9. Leiva, V., Santos, R. A. D., Saulo, H., Marchant, C., & Lio, Y. (2023). Bootstrap control charts for quantiles based on log-symmetric distributions with applications to the monitoring of reliability data. *Quality and Reliability Engineering International*, 39(1), 1–24. <https://doi.org/10.1002/qre.3072>
10. Marchant, C., Leiva, V., & Cysneiros, F. J. A. (2016). A Multivariate Log-Linear Model for Birnbaum-Saunders Distributions. *IEEE Transactions on Reliability*, 65(2), 816–827. <https://doi.org/10.1109/TR.2015.2499964>
11. Mazucheli, J., Alves, B., & Korkmaz, M. Ç. (2023). The Unit-Gompertz Quantile Regression Model for the Bounded Responses. *Mathematica Slovaca*, 73(4), 1039–1054. <https://doi.org/10.1515/ms-2023-0077>
12. Mazucheli, J., Alves, B., Korkmaz, M. Ç., & Leiva, V. (2022). Vasicek Quantile and Mean Regression Models for Bounded Data: New Formulation, Mathematical Derivations, and Numerical Applications. *Mathematics*, 10(9), 1389. <https://doi.org/10.3390/math10091389>

13. Mazucheli, J., Leiva, V., Alves, B., & Menezes, A. F. B. (2021). A New Quantile Regression for Modeling Bounded Data under a Unit Birnbaum–Saunders Distribution with Applications in Medicine and Politics. *Symmetry*, 13(4), 682. <https://doi.org/10.3390/sym13040682>
14. Mazucheli, J., Menezes, A. F. B., & Chakraborty, S. (2019). On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *Journal of Applied Statistics*, 46(4), 700–714. <https://doi.org/10.1080/02664763.2018.1511774>
15. Mazucheli, J., Menezes, A. F. B., Fernandes, L. B., De Oliveira, R. P., & 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. <https://doi.org/10.1080/02664763.2019.1657813>
16. Noufaily, A., & Jones, M. C. (2013). Parametric Quantile Regression Based on the Generalized Gamma Distribution. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 62(5), 723–740. <https://doi.org/10.1111/rssc.12014>
17. Sánchez, L., Leiva, V., Galea, M., & Saulo, H. (2020). Birnbaum-Saunders Quantile Regression Models with Application to Spatial Data. *Mathematics*, 8(6), 1000. <https://doi.org/10.3390/math8061000>
18. Sánchez, L., Leiva, V., Galea, M., & Saulo, H. (2021). Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. *Applied Stochastic Models in Business and Industry*, 37(1), 53–73. <https://doi.org/10.1002/asmb.2556>
19. Sánchez, L., Leiva, V., Saulo, H., Marchant, C., & Sarabia, J. M. (2021). A New Quantile Regression Model and Its Diagnostic Analytics for a Weibull Distributed Response with Applications. *Mathematics*, 9(21), 2768. <https://doi.org/10.3390/math9212768>
R packages:
20. R. Koehler, quantreg: quantile regression, 2021, <https://CRAN.R-project.org/package=quantreg>. R package version 5.86.
21. J. Mazucheli, B. Alves, Ugomquantreg: quantile regression modeling for unit-Gompertz responses, 2021a, R package version 1.0.0.
22. J. Mazucheli, A.F.B. Menezes, unitBSQuantReg: unit-Birnbaum-Saunders quantile regression, 2020, <https://github.com/AndrMenezes/unitBSQuantReg>. R package version 0.1.0.
23. A.F.B. Menezes, Uwquantreg: unit-Weibull quantile regression, 2020, <https://github.com/AndrMenezes/uwquantreg>. R package version 0.1.0.
24. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. <https://www.R-project.org/>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.