International Journal of Mathematical Research
2021 Vol. 10, No. 1, pp. 12-18.
ISSN(e): 2306-2223
ISSN(p): 2311-7427
DOI: 10.18488/journal.24.2021.101.12.18

QUANTILE GENERALIZED ADDITIVE MODEL A ROBUST ALTERNATIVE TO GENERALIZED ADDITIVE MODEL

Nwakuya, Maureen Tobechukwu

Department of Mathematics/Statistics, University of Port Harcourt, Nigeria.
Email: maureen.nwakuya@uniport.edu.ng Tel:+2348033167003

Article History
Received: 5 July 2021
Revised: 19 November 2021
Accepted: 10 December 2021
Published: 27 December 2021

Keywords
Generalized additive model
Quantile generalized additive model
Backfitting algorithm
Spline basis function
Nonparametric regression
Weighted loss function.

ABSTRACT

Nonparametric regression is an approach used when the structure of the relationship between the response and the predictor variable is unknown. It tries to estimate the structure of this relationship since there is no predetermined form. The generalized additive model (GAM) and quantile generalized additive (QGAM) model provides an attractive framework for nonparametric regression. The QGAM focuses on the features of the response beyond the central tendency, while the GAM focuses on the mean response. The analysis was done using gam and qgam packages in R, using data set on live-births, fertility-rate and birth-rate, where, live-birth is the response with fertility-rate and birth-rate as the predictors. The spline basis function was used while selecting the smoothing parameter by marginal loss minimization technique. The result shows that the basis dimension used was sufficient. The QGAM results show the effect of the smooth functions on the response variable at 25th, 50th, 75th and 95th quantiles, while the GAM showed only the effect of the predictors on the mean response. The results also reveal that the QGAM have lower Akaike information criterion (AIC) and Generalized cross-validation (GVC) than the GAM, hence producing a better model. It was also observed that the QGAM and the GAM at the 50th quantile had the same R2adj(77%), meaning that both models were able to explain the same percentage of variation in the models, this we attribute to the fact that mean regression and median regression are approximately the same, hence the observation is in agreement with existing literature. The plots reveal that some of the residuals of the GAM were seen to fall outside the confidence band while in QGAM all the residuals fell within the confidence band producing a better smooth.

Contribution/Originality: This study is one of the very few studies that have investigated quantile generalized additive model as a robust alternative to generalized additive model. In the study of both models, the work revealed through some comparison criteria that QGAM is a better alternative to GAM and also illustrated this through some graphs.

1. INTRODUCTION

The classical approach for estimating a regression function is the parametric regression estimation, but models with additive nonparametric effects offer a valuable dimension reduction device throughout applied statistics. Parametric regression assumes that the structure of the regression function is known and depends only on some parameters, and uses the data to estimate the (unknown) values of these parameters. In linear regression it is assumed that the regression function is a linear combination of the components of the predictor variable for some unknown parameters. The general linear regression model is a form of parametric regression, where the

relationship between the predictor variable 'x' and the response variable 'y' has some predetermined form with the parameterized relationship between them given as, say;

$$Y = X\beta + \epsilon \quad (1)$$

where ϵ is independent and identically distributed random errors, with mean zero, the Equation 1 is known as the regression function of Y on X, where the unknown parameter estimate β(assumed to be linear in the model) are estimated from the data. Furthermore, they are often easy to interpret, for instance in a linear model (when f(x) is a linear function) the absolute value of the coefficient β indicates how much influence a component of X has on the value of Y, and the sign of $\hat{\beta}$ describes the nature of this influence (increasing or decreasing the value of Y). However, parametric estimates have a big drawback. Regardless of the data, a parametric estimate cannot approximate the regression function better than the best function which has the assumed parametric structure [1].

In contrast to parametric regression, the nonparametric regression comes in when the structure of the relationship between the response and the predictor variable is unknown. Nonparametric regression tries to estimate the structure of the relationship between the response and the predictor variable since there is no predetermined form for the relationship between them. The nonparametric regression methods are simply alternative statistical approaches used when some assumptions valid for parametric regression methods are not met. The non-parametric methods make fewer assumptions; they are more flexible, more robust, and applicable to non-quantitative data. The generalized additive model and quantile generalized additive model provides an attractive framework for nonparametric regression. The quantile generalized additive model focuses on the features of the response beyond the central tendency, while the generalized additive model is focused on the mean response. In this work, we intend to show that the quantile generalized additive model is a robust alternative to the generalized Additive models.

2. GENERALIZED ADDITIVE MODEL (REGRESSION SPLINE) (GAM):

A generalized additive model is a nonparametric technique; it is a generalized linear model with a linear predictor having sum smooth functions of the predictor variable. These models assume that the mean of the response variable depends on an additive predictor through a nonlinear link function, Trevor and Robert [2]. Generalized additive models permit the response probability distribution to be any member of the exponential family of distributions. The structural form of the model is given by Equation 2.

$$g(\mu_i) = X\beta + \sum_{j=1}^{p} f_j(x_{ji}) \quad (2)$$

Equation 2 shows the structure form of the Generalized additive model, given that $\mu_i = E(y_i)$ and y_i is the response variable from an exponential family, X is the design matrix of the predictors, β is the corresponding parameter vector, and the fj are smooth functions of the predictors, *x*. The model allows flexible specification of the dependence of the response variable on the predictors, by specifying the model only in terms of 'smooth functions', rather than detailed parametric relationships. Considering a univariate function, we introduce a smooth function of one predictor, given by the form;

$$y_i = f(x_i) + e_i \quad (3)$$

where y_i is a response variable, x_i is a predictor, f is a smooth function and the e_i are i.i.d $\sim N(0,\sigma^2)$ random variables. A regression procedure can be viewed as a method for estimating how the value of y depends on the values of $x_1,\ldots,x_n$. The standard linear regression model assumes the expected value of 'y' has a linear form;

$$E\ (y) = f(x_1,\ldots,x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_n \quad (4)$$

Given a sample of values for y and x, estimates of $\beta_0, \beta_1, \ldots, \beta_p$ are often obtained by the least squares method. The additive model generalizes the linear model by modeling the expected value of y as;

$$E\ (y) = f(x_1,\ldots,x_p)$$

In other to estimate the function *f*, a bases has to be chosen for it, that is to define the space of functions of which f (or a close approximation to it) is an element for some unknown values of β_i, hence;

$$E\ (y) = f(x_1,\ldots,x_p) = s_0 + s_1(x)\beta_1 + \cdots + s_p(x)\beta_p \quad (5)$$

Equation 5 presents the basis function, where $s_j(x), j = 1, \dots, p$ are smooth spline functions in the exponential family, given as its "basis" function. Therefore $s_j(x),$ is the "basis" function and β = [β1 :β2 :…: βp] are the coefficients. The number of "basis" functions depends on the number of inner knots (that is a set of ordered, distinct values of xj) as well as the order of the spline. Specifically, if we let m denote the number of inner knots, the number of 'basis' functions will be given as K = p + 1 +m.

Let's define a quantity;

$$\xi = s_0 + \sum_j^p s_j X_i \tag{6}$$

Equation 6 is the quantity that relates to the mean of the response through a link function, where $s_1(.), \dots, s_p(.)$ are the smooth functions that define the additive component. Therefore we can say that the relationship between the mean μ of the response variable and ξ is defined by a link function $g(\mu) = \xi$. An estimation procedure for additive models known as backfitting was used; it was introduced by Breiman and Friedman [3]. This method allows the component functions of an additive model to be represented using almost any smoothing or modeling technique but the degree of smoothness of a model is hard to integrate into this technique.

The basic idea behind backfitting is to estimate each smooth component of the additive model by iteratively smoothing partial residuals from the additive model, with respect to the predictor(s) that the smooth relates to. The partial residuals relating to the j^{th} smooth term are the residuals resulting from subtracting all the current model term estimates from the response variable, except for the estimate of j^{th} smooth. Almost any smoothing method (and mixtures of methods) can be employed to estimate the smooths. Here is a more formal description of the backfitting algorithm.

1. Set $\hat{s}_0 = \bar{y}$ *and* $\hat{f}_j = linear\ estimate, j = 1, \dots, p$.

2. Obtain the partial residual for j=I,…,p and set $\hat{f}_j$ to be equal to the partial residual.

$$\hat{f}_j = s_j(y - s_0 - \sum_{k \neq j} f_k)$$

3. Continue previous step until the functions stop changing that is until.

$$\max_j \left\| f_j^p - f_j^{p-1} \right\| < \delta$$

Where the δ is approximately 0.

3. QUANTILE GENERALIZED ADDITIVE MODELS (QGAM)

The most popular nonparametric model is the conditional mean regression model. However, compared with a conditional mean function, the conditional quantile regression function, when evaluated at different quantiles, can reveal an entire distributional relationship between the predictor and the response variable. The traditional quantile regression is concerned with the estimation of the τth conditional quantile regression of y for given **x** which often sets a linear model as:

$$Q_y(\tau|X) = X_k^T \beta(\tau) \tag{7}$$

Where X is a vector of predictors, $\beta(\tau)\beta(\tau)$is a vector of the quantile regression coefficients and $y_i y_i$ is a univariate response continuous variable with cdf F(y). To estimate of the coefficients, Koenker and Bassett [4], proposed an L_1-weighted loss function given as;

$$\hat{\beta}(\tau) = \arg \min_{\beta(\tau) \epsilon R^k} \sum_i^n \rho_\tau \left(y_i - X_{ki}^T \beta(\tau) \right), \tau \epsilon (0,1) \tag{8}$$

Equation 8 represents the L_1-weighted loss function, where ρ_τ is a loss function, such that:

$$\rho_\tau(e) = e\left(\tau - \mathbb{I}(e < 0)\right) = \begin{cases} e(\tau - 1), e < 0 \\ e\tau, e \geq 0 \end{cases} \tag{9}$$

Formulating quantile regression as a linear Programming problem, we have the residuals represented as; $e_i = y_i - X_{ki}^T\beta(\tau)$, hence the sum in the minimization problem in Equation 8 can be written as;

$$\sum_i^n \rho_\tau\left(y_i - X_{ki}^T\beta(\tau)\right) = \sum_i^n \rho_\tau(e_i) = \sum_i^n \tau|e_i|\mathbb{I}[e_i \geq 0] + (1-\tau)|e_i|\mathbb{I}[e_i < 0] \quad (10)$$

Equation 10 is an expanded form of L_1-weighted loss function in Equation 8. So that the positive residuals associated with the response which lies above the regression line are assigned weights τ while the negative residuals associated with observed responses below the regression line are assigned weights of $(1-\tau)$. For instance When τ=0.7 each positive residual is weighted 7 times that of a negative residual with weight $1-\tau$=0.3 and so in optimum for every observation above the regression line approximately 7 will be placed below the line. Hence the regression line represents the 0.7 quantile.

Hence the linear program in Equation 3 is analyzed and solved using the standard form;

$$\min_g \; h^T g$$

$$st\ Ag = b,\ g \geq 0$$

To achieve this standard form, g must be positive. To achieve this, residuals are decomposed into positive and negative part using slack variables such that:

$$e_i = u_i - v_i \quad (11)$$

Equation 11 shows the components of the decomposed residual, given that the positive part is; $u_i = \max(0, e_i) = |e_i|\mathbb{I}[e_i \geq 0]$ and the negative part:

$$v_i = \max(0, -e_i) = |e_i|\mathbb{I}[e_i < 0] \quad (12)$$

Equation 12 shows the expanded negative part and positive part of the decomposed residual. The sum of residuals assigned weights by the loss function is then given as;

$$\sum_i^n \rho_\tau(e_i) = \sum_i^n \tau u_i + (1-\tau)v_i = \tau 1_n^T u + (1-\tau)1_n^T v \quad (13)$$

Equation 13 shows the weights assigned by the loss function to the residuals, where 1_n is a n X 1 vector, whose coordinates are all equal to 1. This then results to;

$$\min_{\beta(\tau)\epsilon R^k, u\epsilon R_+^n, v\epsilon R_+^n} \{\tau 1_n^T u + (1-\tau)1_n^T v | y_i = X_i\beta(\tau) + u_i - v_i, i = 1, \ldots, n\} \quad (14)$$

Equation 14 is the minimization form of the problem as given by Koenker-Roger [5].

Quantile generalized additive model (QGAM) assumes that $X_{ki}^T\beta(\tau)$ has an additive structure such that Equation 7 becomes

$$Q_y(\tau|X, f) = X_k^T\beta(\tau) + \sum_{j=1}^{p} f_j(x_{ji})$$

where the p additive terms are fixed smooth functions, defined in terms of spline bases. A marginal smooth functions is given as; say $f_j(x_{ji}) = \sum_j^p s_j X_i\left(\beta_i(\tau)\right)$, where $\beta_i(\tau)\ \beta_i(\tau)$ are unknown coefficients and $s_j X_i$ are known spline basis functions, Fasiolo, et al. [6]. Our aim is to estimate these spline basis functions together with the parameter β at all quantiles. In other to get the estimates we solve;

$$\arg\min_{\beta(\tau)\in R^k, S} \sum_i^n \rho_\tau\left(y_i - X_{ki}^T\beta(\tau) - \sum_j^p s_j X_i\right) + \gamma\|\beta\|_1 + \gamma_j V(\nabla s_j), \tau\epsilon(0,1)$$

where ρ_τ is as defined in Equation 4, $\gamma\|\beta\|_1 = \sum_{k=1}^r |\beta_k|$ and $V(\nabla s_j)$ denotes the total variation of the derivative or gradient of the function s, Koenker-Roger. [7].

4. RESULTS

The analysis was done using a data set on live-births, fertility-rate and birth-rate, where live-birth is the response variable with fertility-rate and bir-thrate as the predictor variables. This analysis was done using the gam and qgam packages in R software.

4.1. Generalized Additive Model results (GAM)

Table-1. Approximate significance of smooth terms.

Smooth terms	edf	Ref.df	F-value	p-value
s(Fertility-Rate)	7.537	8.321	46.34	<2e-16 ***
s(BirthRate)	7.294	8.694	46.09	<2e-16 ***

Table 1 shows that the smooth functions significantly affect the response.

Plots from Generalized Additive Model (GAM)

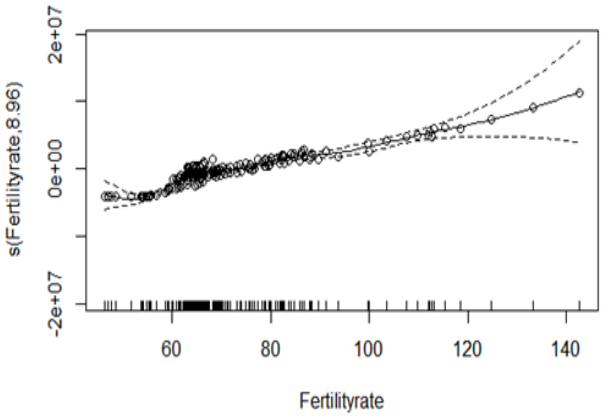

Figure-1. Residual plot for Fertility-rate

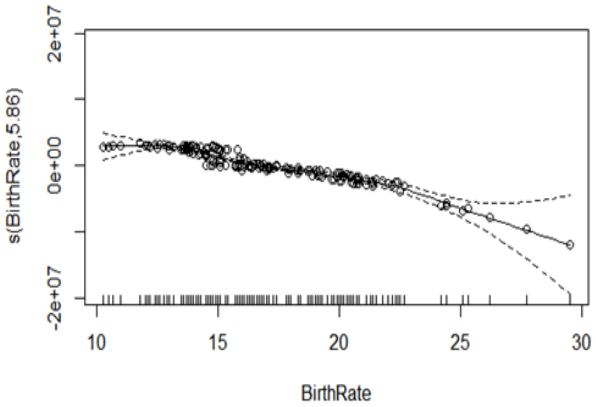

Figure-2. Residual plot for Birth-rate

The Figure 1 & 2 above shows that some of the residual values didn't fall within the confidence band.

4.2. Quantile Generalized Additive Model results (QGAM)

Table-2. Approximate significance of smooth terms.

25th quantile				
Smooth terms	**Edf**	**Ref.df**	**Chi-square**	**p-value**
s(Fertilityrate)	6.441	7.230	441.7	<2e-16 ***
s(BirthRate)	6.773	8.128	515.8	<2e-16 ***
50th quantile				
Smooth terms	Edf	Ref.df	Chi-square	p-value
s(Fertilityrate)	7.126	7.939	559.1	<2e-16 ***
s(BirthRate)	6.863	8.201	580.9	<2e-16 ***
75th quantile				
Smooth terms	Edf	Ref.df	Chi-square	p-value
s(Fertilityrate)	7.364	8.117	1442	<2e-16 ***
s(BirthRate)	10.182	11.689	1485	<2e-16 ***
95th quantile				
Smooth terms	Edf	Ref.df	Chi-square	p-value
s(Fertilityrate)	6.103	6.891	933.4	<2e-16 ***
s(BirthRate)	9.056	10.469	1061.0	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

From the table we can observe that expected degrees of freedom for QGAM are less wiggly in smoothness than that of gam, because their expected degrees of freedom values are smaller except for the smooth function of birthrate for 75th and 95th quantile. We can also see that all the smooth curves for both gam and QGAM show significant changes in the response.

4.3. Plots Form Quantile Additive Model

Figure 3 & 4 shows that in quantile generalized additive models the residual values fall within the confidence band, producing a better smooth than the GAM model.

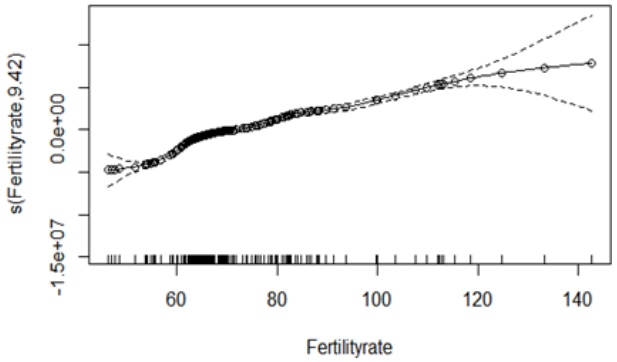

Figure-3. Residual plot for Fertility-rate

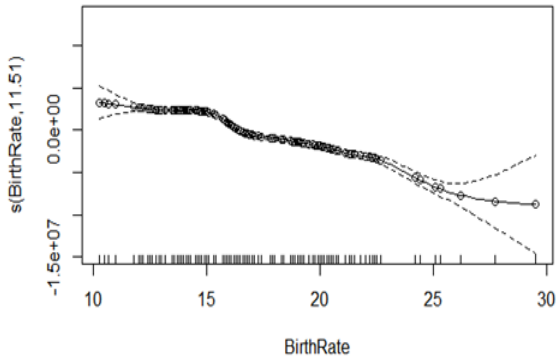

Figure-4. Residual plot for Birth-rate

Table-3. Comparison Criteria.

Models	R-sq.(adj)	Deviance explained	AIC (Akaike information criterion)	GCV (generalised cross-validation)
GAM	0.77	79%	5095.657	4.2931e+11
QGAM(25th quantile)	0.721	79.4%	5087.255	2513.338
QGAM(50th quantile)	0.77	74.4%	5091.31	2516.059
QGAM(75th quantile)	0.726	81.5%	5086.55	2524.178
QGAM(95th quantile)	0.695	96%*	5074.61*	2511.248*

The results show that QGAM models have lowered AIC and GCV's than the GAM model and it's a proof that QGAM model denotes a better performance in comparison with the GAM model. Based on all the models the 95th quantile best fits the model with proportion of deviance explained as 96% and also has the least AIC and GCV denoting the best among all the models (Table 3). It can be said that the outcomes of GAM and QGAM at 50th quantile have shared similar properties in terms of R^2_{adj} which is supposed to be because gam uses response based on mean centered value while 50th quantile uses responses based on the median value which is equivalent.

Table-4. Adequacy of the basis dimension.

Model	Smooth functions	k'	K-index	p-value
GAM	s(Fertilityrate)	19	1.24	1.00
	s(BirthRate)	19	1.01	0.54
QGAM(25th quantile)	s(Fertilityrate)	19	1.05	1.00
	s(BirthRate)	19	0.85	0.42
QGAM (50th quantile)	s(Fertilityrate)	19	1.24	1.00
	s(BirthRate)	19	1.02	0.57
QGAM (75th quantile)	s(Fertilityrate)	19	0.94	0.96
	s(BirthRate)	19	0.84	0.56
QGAM (95th quantile)	s(Fertilityrate)	19	0.45	0.40
	s(BirthRate)	19	0.43	0.15

Table 4 shows significant p-value which indicates the basis dimension chosen is adequate for all the models. Though for the 75th and 95th quantiles there appears to be a missing pattern left in the residuals because the k-index is lower than 1.

5. CONCLUSION

Motivated by the need to show that the quantile generalized additive model is a robust alternative to generalized additive model. The basic framework, outlined above, represents smooth functions in regression models using spline basis function. Selecting the smoothing parameter by marginal loss minimization was done through the fast stable method of Wood, et al. [8]. The result shows that the basis dimension used was sufficient. The results also show that the expected degrees of freedom (edf) for the QGAM were smaller than that of the GAM except for the smooth functions of birthrate at the 75th and 95th quantiles. The comparison criteria in Table 2 reveals that the qgam models have lower AIC and GVC than the gam model, hence a better model. It was observed that the GAM model and the QGAM at the 50th quantile had the same R^2_{adj}(77%), meaning that both models were able to explain the same percentage of variation by the models, this we could attribute to the fact that GAM is based on mean centered value and the QGAM for 50th quantile is based on the median value of the response and in literature mean regression and median regression are approximately the same, hence the observation is in agreement with existing

literature. Also from the plots in fig 1-4, we observe that the residuals of the GAM didn't all fall within the confidence band but for QGAM all the residuals fall within the confidence band producing a better smooth.

Funding: This study received no specific financial support.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

[1] G. Laszlo, k. Michael, A. krzyzak., and W. Harro, "A distribution-free theory of nonparametric regression," ed New York: Springer-Verlay Inc, 2002, pp. 18-30.
[2] H. Trevor and T. Robert, "Generalized additive models," *Statistical Science*, vol. 1, pp. 297-318, 1986.
[3] L. Breiman and J. H. Friedman, "Estimating optimal transformations for multiple regression and correlations," *Journal of the American Statistical Association*, vol. 80, pp. 580–619, 1985.
[4] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, pp. 33-50, 1978.
[5] Koenker-Roger, "Quantile regressions," ed United Kingdom: Cambridge University Press, 2005, pp. 7-14.
[6] M. Fasiolo, S. N. Wood, M. Zaffran, E. Paris, R. Nedellec, and Y. G. Goudé, "Fast calibrated additive quantile regression," *Journal of the American Statistical Association*, pp. 1-11, 2020.
[7] Koenker-Roger., "Additive models for quantile regression: Model selection and confidence bandaids," *Brazilian Journal of Probability and Statistics*, vol. 25, pp. 239–262, 2011.
[8] S. N. Wood, N. Pya, and B. Saefken, "Smoothing parameter and model selection for general smooth models (with discussion)," *Journal of the American Statistical Association*, vol. 111, pp. 1548-1575, 2016.Available at: https://doi.org/10.1080/01621459.2016.1180986.