Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 26, Issue 4 (2024) ISSN$_{Print}$: 1816-9171

Estimating fish body condition using expectiles

Amal Hussien Alwan Rahim Alhamzawi
University of Al-Qadisiyah

Received: 13/8/2024
Accepted: 13/10/2024
Available online: 25 /12 /2024

Corresponding Author : Amal Hussien Alwan

Abstract: In this paper, an expectile regression (ER) approach that includes random effects is suggested in order to account for the dependence between serial observations on the same subject. Specifically, a random intercept expectile regression approach is suggested to estimate the allometric model. We outline the suggested algorithm to make computations easier and provide detailed implementation plans. we use the expectile regression (ER) to estimate the conditional expectiles of the fish weight given length. We examined white bass data in Lake McConaughy, Nebraska, using the suggested methodology. the results show that the suggested approach performs better than the others (the geoadditive expectile regression and the frequentist expectile regression) across different expectiles in terms of the 95% intervals and RMSE.

Introduction:It is well known that fish body weight has an exponential relationship with its length. This exponential relationship is provided by the allometric model that follows.

$$W = \beta_0 L^{\beta}1\ 10^{\varepsilon}, \tag{1}$$

where W represents the weight in grams (g) and L denotes length measured in millimeters (mm), β_0 is the constant of proportionality, β_1 is the allometry coefficient, and ε is the error term. This relationship is a good index in biological and economic analysis of fisheries to assess fish body condition. It shows that the fish body weight has an exponential regression relationship with its length. The regression coefficients (β_0 and β_1) can b estimated by least squares regression using logarithmic (base 10) transformation of weight-length data.

Wege and Anderson (1978) suggested using the relative weight (Wr) to assess fish body condition (Wr = 100 − W/Ws; W refers to the weight of an individual fish, and Ws refers to the expected or average standard weight which is derived for fish of the same species and length). They calculated Ws of fish using the 75th-percentile (i.e., 0.75 quantile) weights from Carlander's method (Carlander et al., 1977), which is commonly used in fisheries biology.
to estimate the weight of fish based on their length. The Ws for fish can vary depending on the species, thus many Ws equations have been proposed over the years. However, a serious challenge with Ws equations is that larger fish within a species commonly have higher condition values than do smaller fish (Wege and Anderson, 1978). Quantile regression can be used to estimate the conditional quantiles of weight of fish as a linear function of the length (Cade et al., 2008; Koenker and Bassett, 1978). Fish weight at length was estimated using quantile-based allometric models, including the 0.75 quantile, by Cade et al. (2008) and Alhamzawi et al. (2011). In Lake McConaughy, Nebraska, the body condition of white bass Morone chrysops was estimated using the quantile regression (QR) model before and after the development of alewives Alosa pseudoharengus between 1980 and 1988 and 1989 to 2004.
A substitute method for QR is expectile regression (ER; Newey and Powell, 1987) which has been widely used in various fields such as demography, economics, genetics, medicine and social science (see, Kneib

(2013), Taylor (2008) and Schnabel and PHC (2009) for an overview). It offers an alternate method to describe the outcome distribution's tails, which is not a generalization of the median as in QR, but rather a generalization of the distribution's center, or mean. According to Kneib (2013), Sobotka et al. (2013), and Waltrup et al. (2015), it is simple to convert to calculate QR.Compared to QR, ER is fitted using quadratic optimization (Newey and Powell, 1987) while QR is fitted using linear programming (Koenker, 2004; Koenker and d'Orey, 1987). Quantile and expectile regressions provide a more flexible and comprehensive understanding of the conditional distribution of the response compared to traditional regression models (Foroni, B., Merlo, L., & Petrella, L., 2023).

Regression analysis is a statistical modeling technique that examines the relationship between variables using optimization loss functions (Xu, Q. F., Ding, X. H., Jiang, C. X., Yu, K. M., & Shi, L. ,2021)

This study aims to estimate the expectiles of fish data's length-weight relationship using an allometric model's expectile analysis.

Specifically, we use the expectile regression (ER) (Newey and Powell, 1987) to estimate the conditional expectiles of the fish weight given length fish. In other words, it allows you to model different parts of the distribution of the response variable. The study utilized our methodology to analyze data on white bass Morone chrysops from Lake McConaughy, Nebraska.

Methods:

Clustered data with random effects:

For the white bass data, we highlight in Figure (1) how weights vary with length across time. The outcome variable calculates each person's weight (g) based on their classification according to years. The data from the same year is highly likely to be statistically associated rather than independent since the observations are made in clusters, or years. Using a linear model to examine the white bass data in this instance is improper.

In this paper, we consider a clustered data of white bass Morone chrysops were conducted yearly surveys of Lake McConaughy, Nebraska, between 1989 and 2004. Captured fish were weighed to the closest gram and measured to the closest millimeter.Given clustered data $\{W_{it}, L_{it};\ i = 1, \cdots, N,\ t = 1, \cdots, n_i\}$, model (1) can be written as (Alhamzawi et al., 2011)

$$W_{it}=\beta_0 L_{it}^{\beta_1}\, 10^{(u_i\, \varepsilon_{it})}, \qquad i=1,\ldots,N;\ t=1,\ldots,n_i. \qquad (2)$$

The tth weight in cluster i is W_{it}, while the tth total length in cluster i is L_{it}., β_0 is the ith cluster's position shift random effect, u_i is the allometry coefficient, and is the common constant for all clusters. and ε_{it} is the error term.

A logarithmic transformation of (2) leads to a linear relationship

$$W_{it}^{*}=\beta_0^{*}+\beta_1 L_{it}^{*}+\mu_i+\varepsilon_{it} \qquad (3)$$

where $W_{it}^{*} = \log_{10} W_{it}$, $\beta_0^{*} = \log_{10} \beta_0$ and $L_{it}^{*} = \log_{10} L_{it}$. Let $0 < \tau < 1$ is the expectile level. Under $0 < \tau < 1$, we assume that the conditional expectile of W_{it}^{*}, $E_{W_{it}^{*}|u_i,L_{it}^{*}}$, is given by $\beta_0^{*} + \beta_1 L_{it}^{*} + u_i$. We presume that the τ th expectile of the model's mistake ε_{it} is zero. That is,

$$E_{W_{it}^{*}|\mu_i,L_{it}^{*}}(\tau)=\beta_0^{*}+\beta_1 \log_{10} L_{it}+u_i. \qquad (4)$$

The minimization problem can be solved to estimate the expectile regression coefficients in (4).

$$\sum_{i=1}^{N} \sum_{=1} \tau\ ?W_{it}^{*}-\beta_0^{*}-\beta_1 L_{it}^{*}-u_i\ . \qquad (5)$$

where ρ_τ is the asymmetric quadratic loss (Newey and Powell, 1987) which is defined by $\rho_\tau = \varepsilon^2.|\tau - I(\varepsilon < 0)|$, where $I(\bullet)$ is used as an indicator.. When $\tau = 1/2$, (5) reduces to

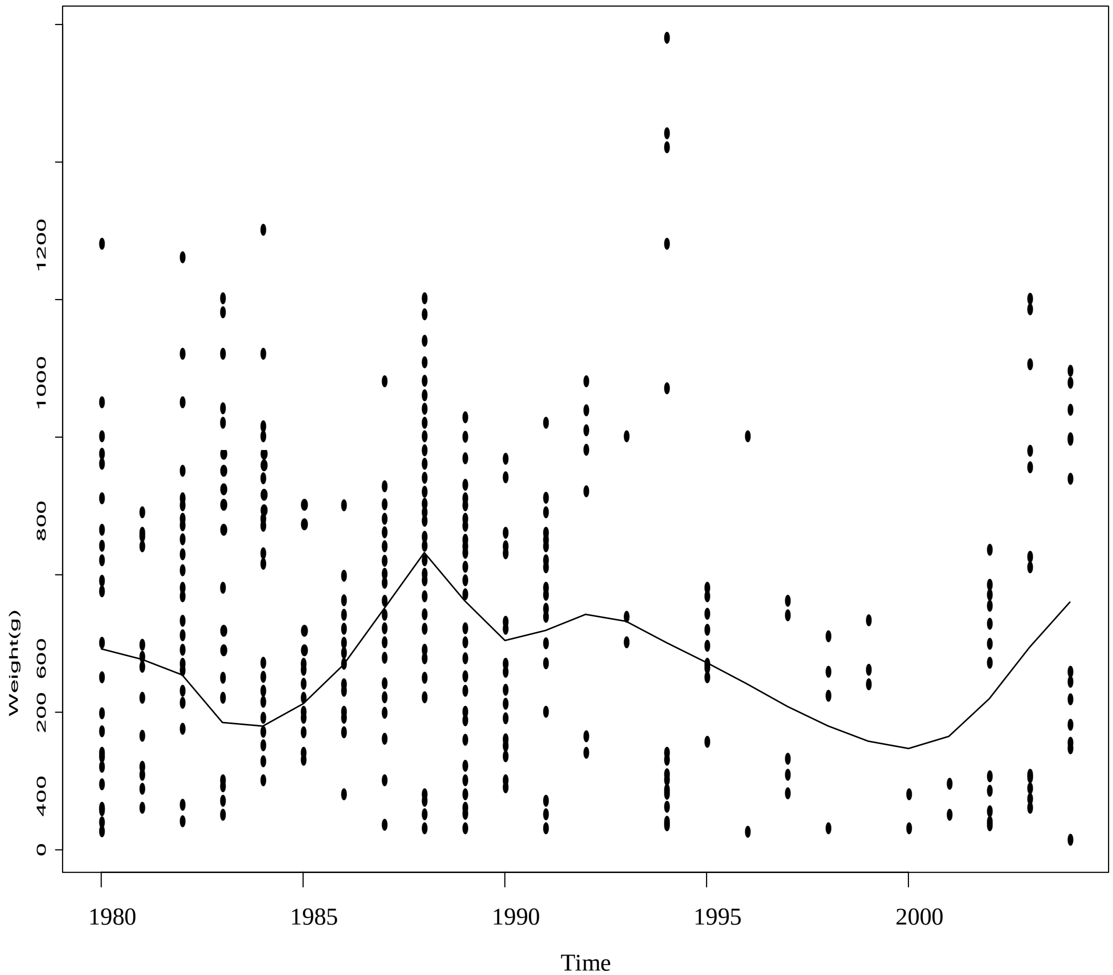

Figure 1: Data on white bass Morone Chrysops are shown in Figure 1 as a scatter plot. The variation in weight in grams over time is displayed as a smooth fitted line.

quadratic loss, which is the standard minimization problem of the intercept model. For all other values of τ , the minimization problem in (5) is asymmetric quadratic loss. We assume that $\{W_{i\,t}^{*};\ i = 1, \cdots, N,\ \ t = 1, \cdots, n_i\}$, conditionally on u_i, are dispersed independently using the asymmetric normal distribution (AND) , $W_{i\,t}^{*} \sim$ AND(μ_{it}, σ^2, τ), where $\tau \in(0, 1)$ is the skew parameter, σ^2 is the scale parameter, μ_{it} is the location parameter. Let $\mu_{it} = \beta 0* + \beta_1 L{*}it + u_i$ is the linear predictor of the τ th expectile. The density of AND is given by

$$f(W_{it}^{*} \backslash L_{it}^{*}, \beta_0^{*}, \beta_1 \sigma^2) = \frac{2}{\sqrt{\sigma 2\pi}}\left(\sqrt{\frac{/}{-\tau}}\ 1/\tau\right) exp\left(-\frac{1}{2\sigma^2}\ (W_{it}^{*} - u_{it})^2 |\tau - I(W_{it}^{*} - u_{it} < 0)|\right). \quad (6)$$

1.1 Random effect distribution

We believe that the location shift random effect u_i are mutually identically distributed and independent based on a density f_u characterized by a parameter θ, where θ is dependent on the expectile level (*i.e,* $\theta(\tau$)). Under this assumption, we assume that u_i has an asymmetric normal distribution, $u_i \sim$ AND(0, θ, τ), where $\theta$is the scale parameter. i.e,

$$\mathrm{f}(\mu_{i\,|}\ \theta) = \frac{2}{\sqrt{\theta\pi}}\left(\sqrt{\frac{1}{1-\tau} + \quad \frac{1}{\tau}}\right) exp\left(-\frac{1}{2\theta}\ (u_i^2 |\tau - I(u_i\ < 0)|\right) \quad (7)$$

Let $\mathbf{W} = (W_{i1}^{*}, \cdots, W_{ini}^{*})'$, $\mathbf{W}^{*} = (W_i^{*}, \cdots, W_N^{*})'$ and $\mathbf{u}_i = (u_1, \cdots, u_N)'$. The complete-data density of ($\mathbf{W}^{*}$, $\mathbf{u}$) is given by

$$f(\mathbf{W}^{*}, \mathbf{u}|\beta_0^{*}, \beta_1, \sigma^2, \theta) = \prod_{i=1}^{N} f(W_i^{*}, \beta_0^{*}, \beta_1, \sigma^2, u_i) f(u_i|\theta), \quad (8)$$

and the marginal distribution of the random effect is given by

$$f(u_i|W_i^{*}, \beta_0^{*}, \beta_1, \sigma^2) \propto f(W_i^{*}, \beta_0^{*}|, \beta_1, \sigma^2, u_i) f(u_i|\theta) \quad (9)$$

$$= \frac{4}{\sqrt{\theta\sigma^2\pi}}\sqrt{\frac{1}{1-\tau} + \frac{1}{\tau}}\ \exp\left(-\frac{(W_{it}^{*} - u_{it})^2}{2\sigma^2}\ |\tau - I(W_{it}^{*} - u_{it} < 0)| - \frac{u_i^2}{2\theta}\ |\tau - I(u_i\ < 0)|\right). \quad (10)$$

1.2 Estimation

Following Geraci and Bottai (2007), we obtain the maximum likelihood estimate (MLE) of β_{0^*}, β_1, σ^2, and θ by applying the following iterative procedure:

Input: (W_{it}^*, L_{it}^*) *for* $i=1,\ldots,N$ *and* $t+1,\ldots,n_i$

Initialize: $(\beta_0^*, \beta_1, \sigma^2, \theta)$

For $t = 1,\ldots,(t_{max} + t_{burn\text{-}in})$ **do**

1. Sample K_i observations $v_i = (v_{i1}, \ldots, v_{ik1})'$ from
   $f(\mu_i|\boldsymbol{W}^*_i, \beta_0^*, \beta_1, \sigma^2, \theta) \propto f(\boldsymbol{W}^*_i, \beta_0^*, \beta_1, \sigma^2, \mu_i) f(\mu_i|\theta)$

2. Solve

   $\sum_{i=1}^{N} \frac{/}{} \sum_{m=1}^{Ki} \quad \sum_{=1^\wedge}^{i} \quad (\mathbf{W}^*_{it} - \beta^*_0 - \beta_1 L_{it}^* - v_{im})$, where $\rho_\tau(\varepsilon) = \varepsilon^2 .|\tau - I(\varepsilon < 0)|$.

3. Calculate σ^2,where

   $$\sigma^2 = \frac{1}{\sum_{m=1}^{Ki} \quad \sum_{=1^\wedge}^{i}} \sum_{i=1}^{N} \quad \sum_{=1^\wedge}^{i} \quad \sum_{1^\wedge n}^{\rho} \quad ((\boldsymbol{W}_{it}^* - \beta_0^* - \beta_1 L_{it}^* - v_{im})$$

4. Calculate the MLE of θ ,where

   $\theta = 1N^{1i=1Ni} i=1N=1^\wedge ii$

End for

Figure 2: MLE of the parameters β_{0^*}, β_1, σ^2, and θ.

Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{\text{Online}}$: 2312-9883
QJAE, Volume 26, Issue 4 (2024) ISSN$_{\text{Print}}$: 1816-9171

2 Analysis of the white bass data

The length–weight relationship of white bass in Lake McConaughy, Nebraska, was estimated using expectile regression. Figure 3 illustrates the correlation between the white bass fish's weight and total length measurements. Porath et al. published information about the study area and fish sample (2003). Multiple authors examined this data collection. For instance, Alhamzawi et al. (2011) utilized Regression using Bayesian quantiles for longitudinal data to estimate the regression parameters for this data set, while Cade et al. (2008) quantile regression to compare the white bass's physical condition Sander vitreus before (1980–1988) and after (1989–2004).

. Figure 4 demonstrates the connections between the white bass's weight and total length (TL). data, which shows clearly that some data are outliers.

It is evident from Figure 4's scatter plot that not every observation falls precisely on the estimated fitted line. This implies that the observed weight values may differ from the projected weight values for a given length. The residual errors, shown by a vertical red line, are the differences between the observed and anticipated weights. Figure (5) shows the confidence bands for the fitted curve using the qplot() function in the ggplot2 package. Figure (6) shows the most extreme observations in the data set. These extreme observations are potentially problematic. The top left penal (the residual plot) shows no fitted pattern. That is, the red line should be approximately horizontal at zero. The presence of a pattern may indicate a problem with some aspect of the linear model. The below left penal (the residual plot) shows the variability (variances) of the residual points increases with the value of the fitted observations, suggesting a heteroscedasticity problem in the data. The QQ plot of residuals in he top right penal shows that.

Since the majority of the points roughly fall around the reference line, normalcy might be assumed.The below right penal (the residual plot) shows the top 3 most extreme points (133, 135 and 301), with a standardized residuals exceed 3 standard deviations which means these points are outliers. These outlier points encourage us to use the expectile regression.

Three models are compared: The geoadditive expectile regression, sometimes known as "GER," is performed using the R "expectreg" package (R Core Team, 2024) and is a frequentist expectile regression., known as “FER” and the suggest method, referred to as “RER”. Methods are compared based confidence intervals and The definition of root mean squared error (RMSE) is

$$\mathrm{RMSE}(\theta) = \sqrt{\frac{\sum_{i=1}^{N} \sum_{=1^\wedge}^{-} Expectile\_W_{ij}(\theta|L_{ij}) - \widehat{Expectile}W_{ij}(\theta|L_{ij})^2}{\sum_{i=1}^{N} -}},$$

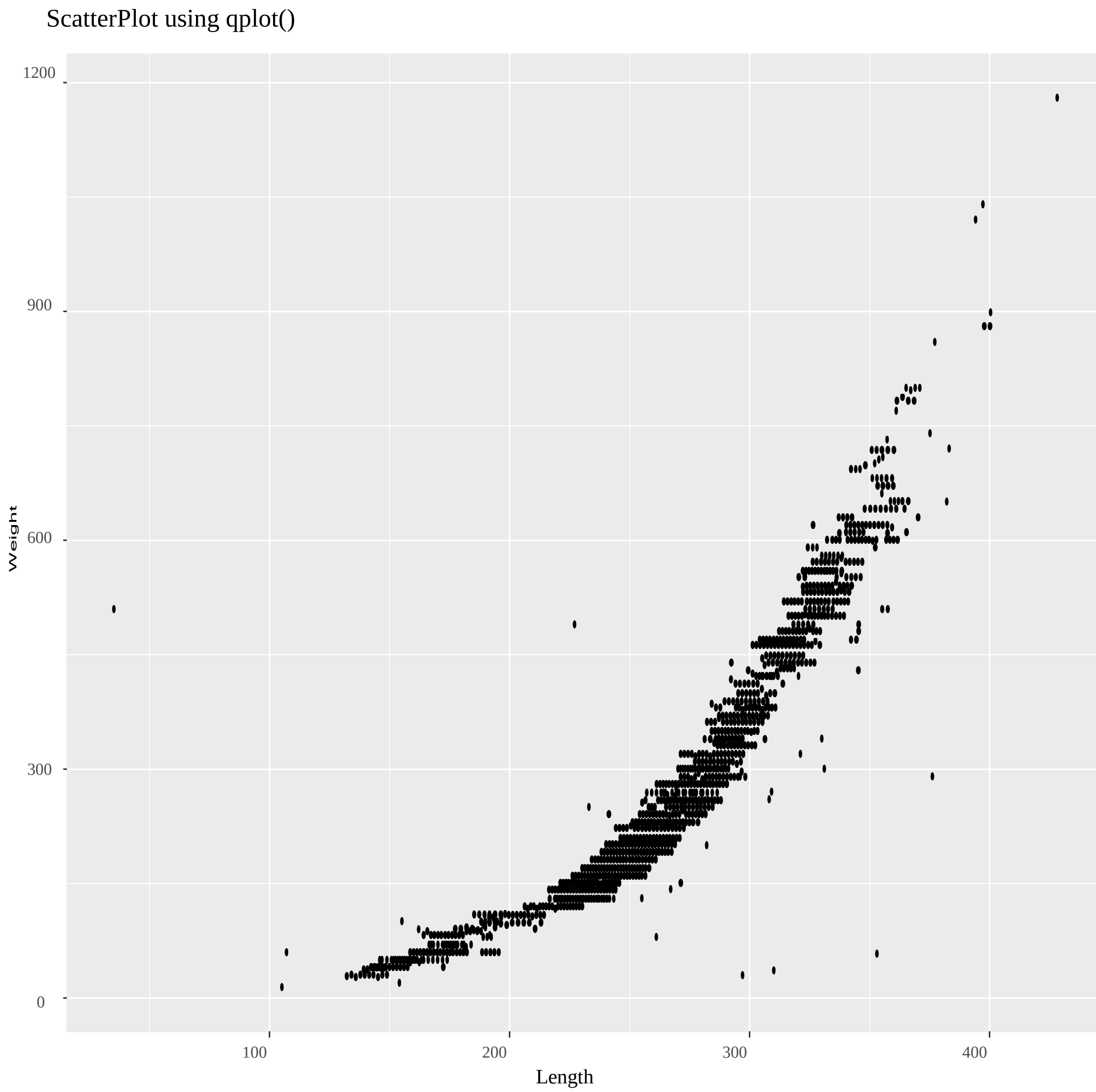

Figure 3: The relationship between weight and total length observations of the white bass fish.

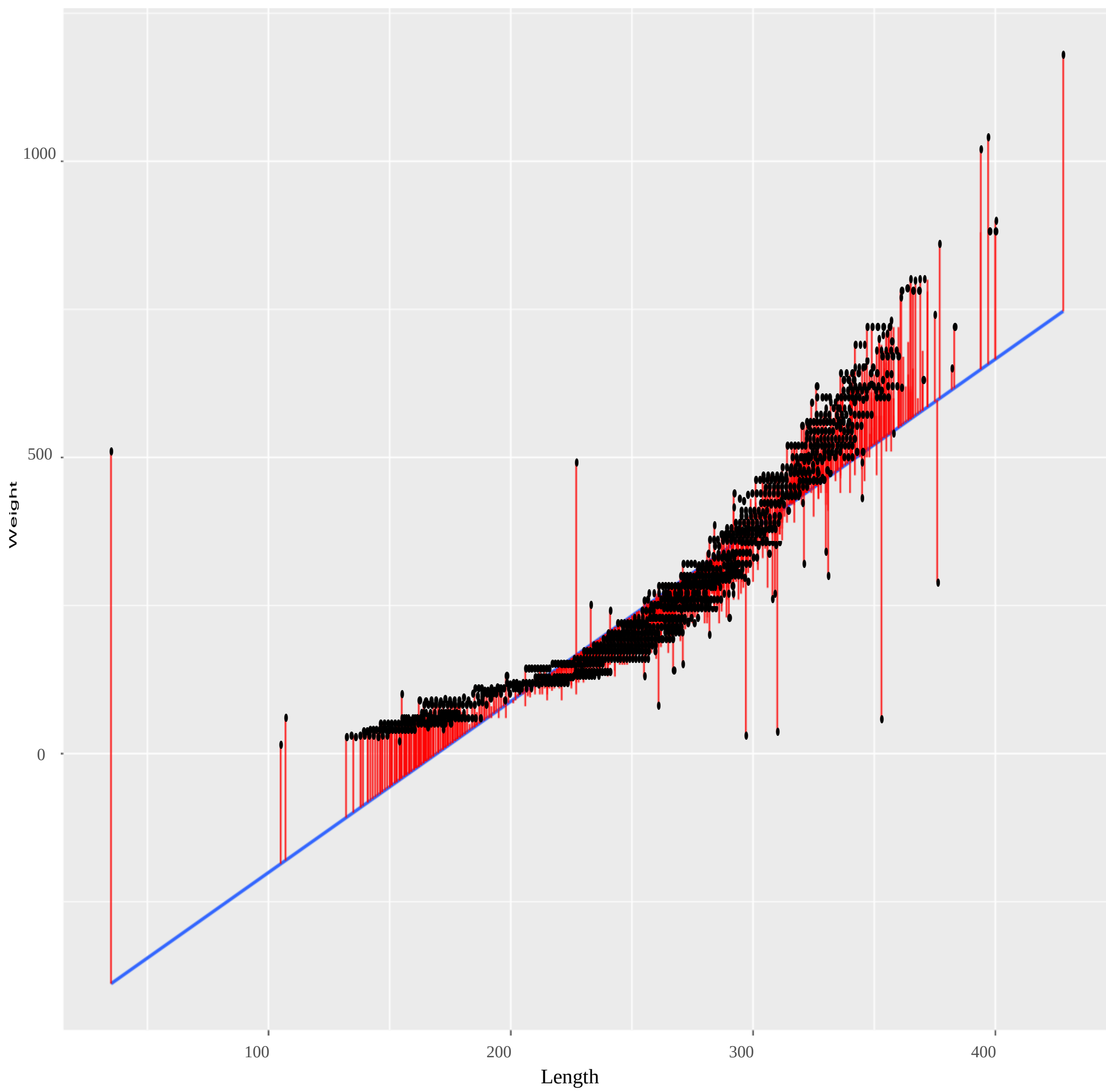

Figure 4: The fitted (or predicted) values for the relationship between weight and total length observations of the white bass fish.

$ISSN_{Online}$: 2312-9883
$ISSN_{Print}$: 1816-9171

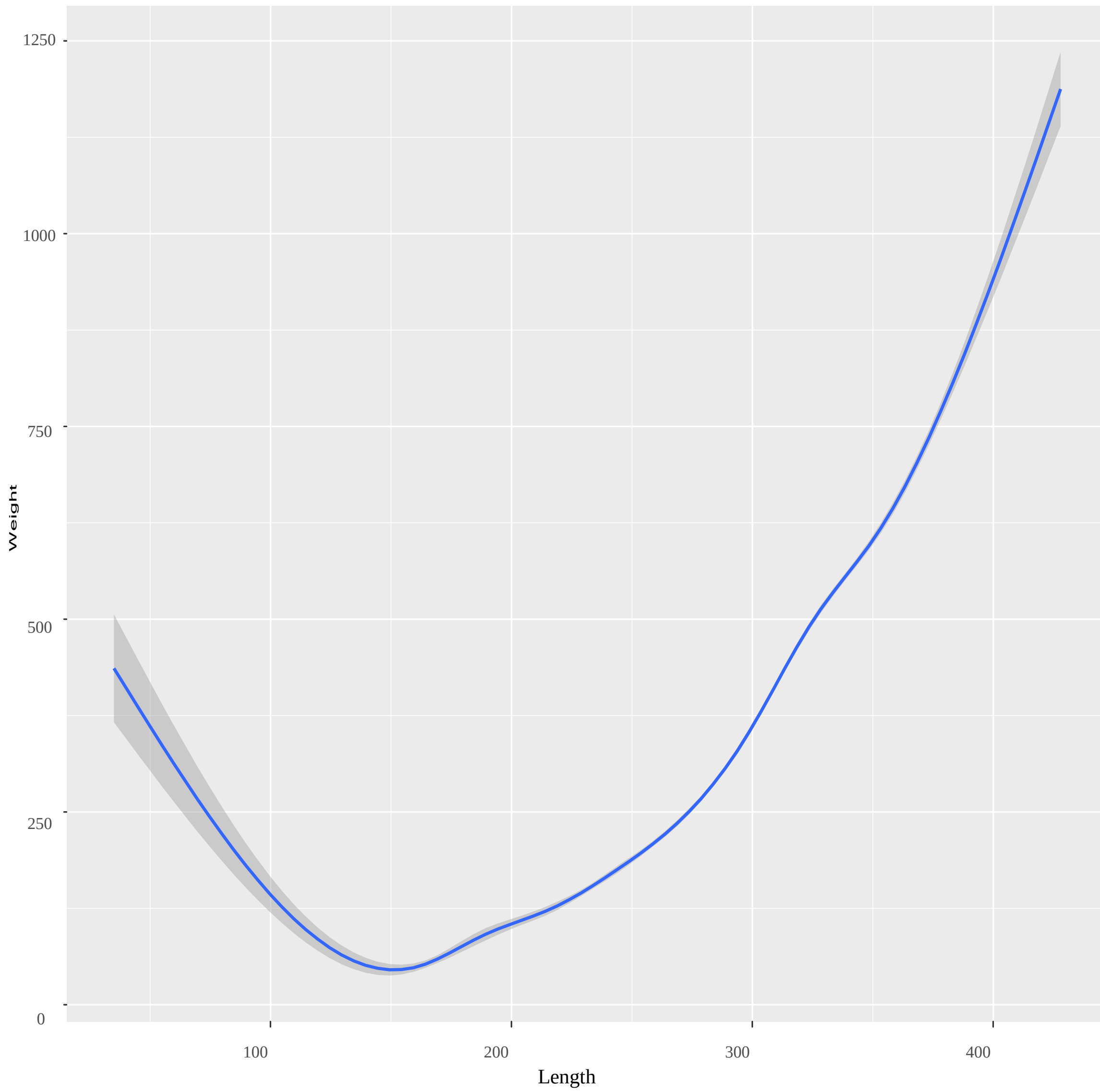

Figure 5: Confidence bands for the fitted curve using the qplot() function in the ggplot2 package

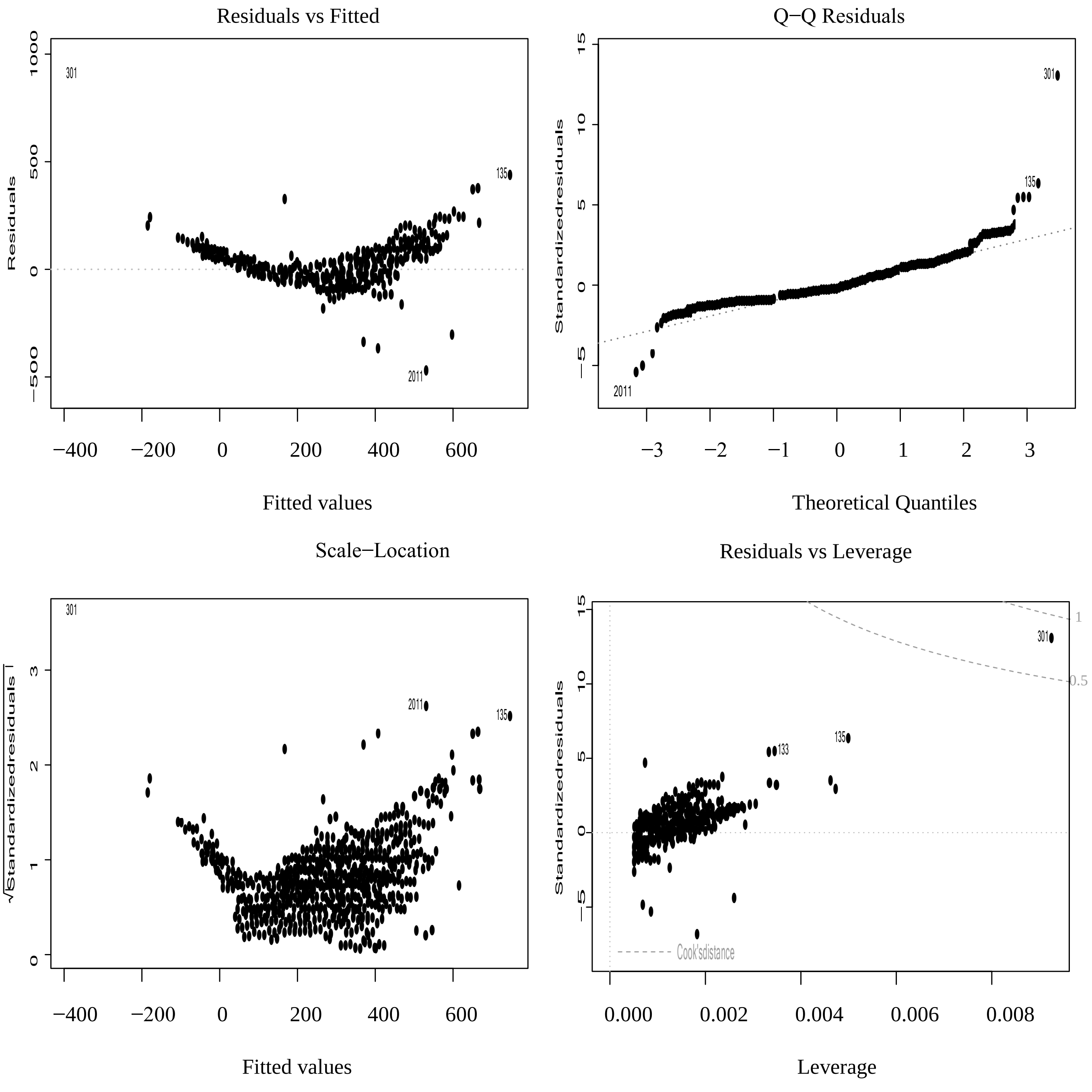

Figure 6: Regression diagnostics plots.

where $Ex\hat{}pectile_{Wij}$ ($\theta|L_ij$) is the estimated value of true expectile *ExpectileWij* ($\theta|L_ij$) of W_ij given L_ij. In general, the smaller value of RMSE is, the better the model performs.

Table. The white bass data's root mean squared errors (RMSE) and 1:95% intervals
. The suggested method is compared with two other approaches: the geoadditive expectile regression and the frequentist expectile regression.

Model	Expectile	β_0	β_1	RMSE
GER	0.10	(-5.875, -5.842)	(3.306, 3.329)	0.089
FER	0.10	(-6.219, -6.116)	(3.319, 3.411)	0.096
RER	0.10	(-5.837, -5.821)	(3.311, 3.321)	0.078
GER	0.25	(-5.950, -5.662)	(3.199, 3.315)	0.073
FER	0.25	(-5.961, -5.671)	(3.201, 3.428)	0.068
RER	0.25	(-5.802, -5.677)	(3.087, 3.299)	0.063
GER	0.50	(-5.819, -5.637)	(3.257, 3.281)	0.057
FER	0.50	(-5.821, -5.677)	(3.244, 3.290)	0.069
RER	0.50	(-5.801, -5.738)	(3.260, 3.279)	0.049
GER	0.75	(-5.634, -5.611)	(3.227, 3.285)	0.077
FER	0.75	(-5.418, -5.400)	(3.241, 3.292)	0.064
RER	0.75	(-5.618, -5.534)	(3.230, 3.275)	0.062
GER	0.90	(-5.519, -5.429)	(3.223, 3.388)	0.091
FER	0.90	(-5.521, -5.444)	(3.231, 3.376))	0.087
RER	0.90	(-5.618, -5.599)	(3.219, 3.279)	0.082

At lower expectiles ($\tau = 0.10$), the estimated parameters of β0 and β1 are roughly $10^{-5.866}$ and 3.318, respectively, while at higher expectiles ($\tau = 0.90$), they are approximately $10^{-5.566}$ and 3.306. The 95% intervals for the three approaches at the five distinct expectiles—0.10, 0.25, 0.50, 0.75, and 0.90—are summarized in Table 1. Evidently, the 95% intervals are typically somewhat broader than the recommended approach (RER). For instance, the interval width of the recommended approach for log10β0 and β1 is 0.019 and 0.060, respectively, at higher expectiles $\tau = 0.90$, compared with 0.077 and 0.145 for GER and 0.090 and 0.165 for the FER method, respectively. The recommended method's interval widths for log10β0 and β1 at lower expectiles, $\tau = 0.10$, are 0.016 and 0.010, respectively, while GER and FER methods have interval widths of 0.033 and 0.023 and 0.103 and 0.092, respectively. The proposed method's interval width in the center of the distribution ($\tau = 0.50$) for log10β0 and β1 is 0.063 and 0.019, respectively, whereas the GER method's interval width is 0.182 and 0.024, and the FER method's is 0.144 and 0.046.

. Table 1 presents a summary of the findings of each model's RMSE comparison. It is obvious that these approaches produce essentially the same results. But overall, the comparison of root mean squared error (RMSE) indicates that the recommended approach is more effective than the geoadditive expectile regression(GER) and the frequentist expectile regression (RER).

3 Discussion

With the main objective method calculating models with random intercepts, we have proposed an expectile regression approach for the allometric model estimation in this study. We have shown the recommended approach to make the computation easier and provide thorough implementation strategies. We analyzed white bass data from Lake McConaughy, Nebraska, using the recommended methodology. According to the data, the recommended method outperforms the other(the geoadditive expectile regression and frequentist expectile regression) when it comes to 95%

intervals and RMSE for various expectiles.

References

Alhamzawi, R., K. Yu, V. Vinciotti, and A. Tucker (2011). Prior elicitation for mixed quantile regression with an allometric model. Environmetrics 22 (7), 911–920.

Cade, B. S., J. W. Terrell, and M. T. Porath (2008). Estimating fish body condition with quantile regression. North American Journal of Fisheries Management 28, 349–359.

Carlander, K. D. et al. (1977). Handbook of freshwater fishery biology: Volume 2. Iowa State University Press.

Foroni, B., Merlo, L., & Petrella, L. (2023). Quantile and expectile copula-based hidden Markov regression models for the analysis of the cryptocurrency market. arXiv preprint arXiv:2307.06400.

Geraci, M. and M. Bottai (2007). Quantile regression for longitudinal data using the asym- metric Laplace distribution. Biostatistics 8, 140–154.

Kneib, T. (2013). Beyond mean regression. Statistical Modelling 13 (4), 275–303.

Koenker, R. (2004). Quantile regression for longitudinal data. Journal of Multivariate 91, 74–89.

Koenker, R. and G. Bassett (1978). Regression quantiles. Econometrica: Journal of the Econometric Society 46 (1), 33–50.

Koenker, R. W. and V. d'Orey (1987). Algorithm as 229: Computing regression quantiles.
Journal of the Royal Statistical Society. Series C (Applied Statistics) 36 (3), 383–393.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing.
Econometrica: Journal of the Econometric Society 55 (4), 819–847.

Porath, M. T., E. J. Peters, and D. L. Eichner (2003). Impact of alewife introduction on walleye and white bass condition in lake mcconaughy, nebraska, 1980-1995. North American Journal of Fisheries Management 23, 1050–1055.

R Core Team (2024). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Schnabel, S. K. and E. PHC (2009). Non-crossing smooth expectile curves. In Proceedings of the 24th International Workshop on Statistical Modelling. Ithaca, NY, USA, pp. 330–36.

Sobotka, F., G. Kauermann, L. Schulze Waltrup, and T. Kneib (2013). On confidence intervals for semiparametric expectile regression. Statistics and Computing 23, 135–148.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. Journal of financial econometrics 6 (2), 231–252.

Waltrup, L. S., F. Sobotka, T. Kneib, and G. Kauermann (2015). Expectile and quantile regression—david and goliath? Statistical Modelling 15 (5), 433–456.

Wege, G. and R. Anderson (1978). Relative weight (wr): a new index of condition for large- mouth bass. New approaches to the management of small impoundments. American Fish- eries Society, North Central Division, Special Publication 5, 79–91.

Xu, Q. F., Ding, X. H., Jiang, C. X., Yu, K. M., & Shi, L. (2021). An elastic-net penalized expectile regression with applications. Journal of Applied Statistics, 48(12), 2205-2230.