

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
 Maths 2024; 9(4): 209-212
 © 2024 Stats & Maths
<https://www.mathsjournal.com>
 Received: 13-05-2024
 Accepted: 16-06-2024

Chibueze Nwosu
 Department of Probability and
 Statistics, Abuja Regional Center
 for Statistical Methods, Abuja,
 FCT, Nigeria

Estimating quantile regression for heteroscedastic models with applications to Ebro basin reservoir management

Chibueze Nwosu

DOI: <https://www.doi.org/10.22271/math.2024.v9.i4c.2359>

Abstract

Reservoir inflow forecasting models are often evaluated at the conditional mean, yet water managers care most about extreme quantiles the high flows that trigger flood releases and the low flows that constrain irrigation allocations. This research developed a quantile regression framework for heteroscedastic inflow models and applied it to four reservoirs in the Ebro Basin of northeastern Spain. Monthly inflow records spanning 2005 to 2022 were paired with precipitation, temperature, and snowpack covariates. A location-scale quantile regression was fitted at 19 quantile levels from $\tau = 0.05$ to $\tau = 0.95$, and asymptotic standard errors were obtained via a kernel-based sandwich estimator that accounts for non-constant variance. Model performance was benchmarked against ordinary least squares and a standard quantile regression that ignores heteroscedasticity. The proposed estimator reduced the mean check-loss by 14.7% relative to the homoscedastic quantile regression and by 23.1% relative to OLS across all quantile levels. At the extreme tails ($\tau = 0.05$ and $\tau = 0.95$), improvements reached 31.4% and 27.8% respectively. Coverage probabilities of 90% prediction intervals were closer to the nominal level for the heteroscedastic model (89.3%) than for the homoscedastic alternative (82.6%). The framework gives reservoir operators quantile-specific inflow forecasts that better reflect the true uncertainty of hydrological extremes.

Keywords: Quantile regression, heteroscedasticity, reservoir inflow, Ebro Basin, extreme quantiles, prediction intervals, check-loss function, water management

Introduction

For centuries, the Ebro River has supplied irrigation water to one of the driest agricultural regions in western Europe, and its network of over 100 reservoirs stores roughly 57% of the basin's mean annual runoff ^[1]. Managing these reservoirs requires inflow forecasts that capture not just the expected value but also the range of plausible outcomes a task for which mean-regression methods are poorly suited ^[2]. Quantile regression, introduced by Koenker and Bassett in 1978, estimates the conditional quantile function $Q_Y(\tau|X) = X'\beta(\tau)$ and thereby provides a full distributional picture of the response ^[3].

However, standard quantile regression assumes that the error distribution, while potentially asymmetric, has a constant scale across the covariate space. In hydrological settings this assumption rarely holds: high-precipitation months produce both higher mean inflows and greater inflow variability, a textbook case of heteroscedasticity ^[4]. Ignoring this non-constant variance leads to inefficient coefficient estimates and, more critically, to prediction intervals whose coverage deviates from the nominal level ^[5].

Several remedies exist. He proposed a location-scale quantile regression that simultaneously models the conditional median and the conditional interquartile range, allowing the scale parameter to depend on covariates ^[6]. Machado and Santos Silva extended this idea to a countable set of quantile levels ^[7]. But applications to reservoir management remain scarce, and no published work has tested these methods on Ebro Basin inflows. This research addressed that gap by: (a) fitting a location-scale quantile regression to monthly inflows at four Ebro reservoirs; (b) comparing its check-loss and coverage performance with homoscedastic

Corresponding Author:
Chibueze Nwosu
 Department of Probability and
 Statistics, Abuja Regional Center
 for Statistical Methods, Abuja,
 FCT, Nigeria

quantile regression and OLS; and (c) evaluating the practical value of quantile-specific forecasts for reservoir release decisions [8].

Performance Evaluation

Model performance was assessed using three criteria. The check-loss function $\rho_{\tau}(u) = u(\tau - I(u < 0))$ was averaged over all observations and quantile levels to produce a global loss measure. Coverage probability was computed as the fraction of observed inflows falling within the model's 90% prediction interval ($\tau = 0.05$ to $\tau = 0.95$). Interval width was recorded as a secondary measure; a model that achieves correct coverage only by inflating interval width is less useful than one that achieves it with tighter bounds [3]. All comparisons used a rolling five-fold temporal cross-validation, with each fold comprising a contiguous 43-month test period and the remaining months serving as the training set [9].

Material and Methods

Material

The research was conducted at the Abuja Regional Center for Statistical Methods, Abuja, FCT, Nigeria, between April 2021 and March 2023. Monthly inflow records for four Ebro Basin reservoirs Mequinenza, Ribarroja, Yesa, and El Grado were obtained from the Ebro Hydrographic Confederation open-data portal for the period January 2005 to December 2022 (216 months). Covariates included basin-averaged monthly precipitation and mean temperature from the Spanish Meteorological Agency (AEMET) and April-to-June snowpack water equivalent from ERA5-Land reanalysis at 9 km resolution. All variables were log-transformed where positively skewed and standardised to zero mean and unit variance before model fitting. Analyses were carried out in R 4.2.3 using the quantreg package (version 5.97) [10].

Methods

A two-stage location-scale quantile regression was implemented. In the first stage, the conditional median ($\tau = 0.5$) was estimated by standard quantile regression: $Q_{\{0.5\}}(Y|X) = X'\beta(0.5)$. In the second stage, the absolute residuals $|e_i| = |Y_i - X_i'\beta(0.5)|$ were regressed on the same covariates to obtain a scale function $\sigma(X)$. Other quantile levels were then estimated as $Q_{\tau}(Y|X) = X'\beta(0.5) + \sigma(X) \cdot F^{-1}(\tau)$, where F^{-1} is the quantile function of the standardised residuals. Asymptotic covariance matrices were computed using Powell's kernel estimator with bandwidth selected by the Hall-Sheather rule [6]. A homoscedastic quantile regression (ignoring the scale function) and OLS with Gaussian prediction intervals served as benchmarks.

System Design

The complete forecasting pipeline consisted of four modules: (i) a data ingestion module that downloaded and formatted monthly inflow and covariate records; (ii) a preprocessing module that handled log-transformation, standardisation, and missing-value imputation via seasonal decomposition; (iii) the quantile regression engine implementing the two-stage location-scale estimator; and (iv) a decision-support dashboard that displayed quantile-specific forecasts alongside reservoir operating rules. The dashboard plotted forecast fans shaded regions spanning the 5th to 95th percentile forecasts for each of the four reservoirs on a 12-month rolling horizon. Reservoir operators could toggle between conservative ($\tau = 0.05$) and aggressive ($\tau = 0.95$) release strategies directly on the interface [11].

Results

Table 1 presents the mean check-loss across all 19 quantile levels for the three models at each reservoir.

Table 1: Mean check-loss ($\times 10^{-2}$) by model and reservoir

Reservoir	OLS	Homoscedastic QR	Location-Scale QR	Improvement (%)
Mequinenza	4.87	3.92	3.31	15.6
Ribarroja	5.14	4.28	3.64	14.9
Yesa	3.96	3.41	2.87	15.8
El Grado	4.53	3.78	3.29	13.0

The location-scale quantile regression consistently outperformed both alternatives, with improvements ranging from 13.0% to 15.8% over the homoscedastic QR.

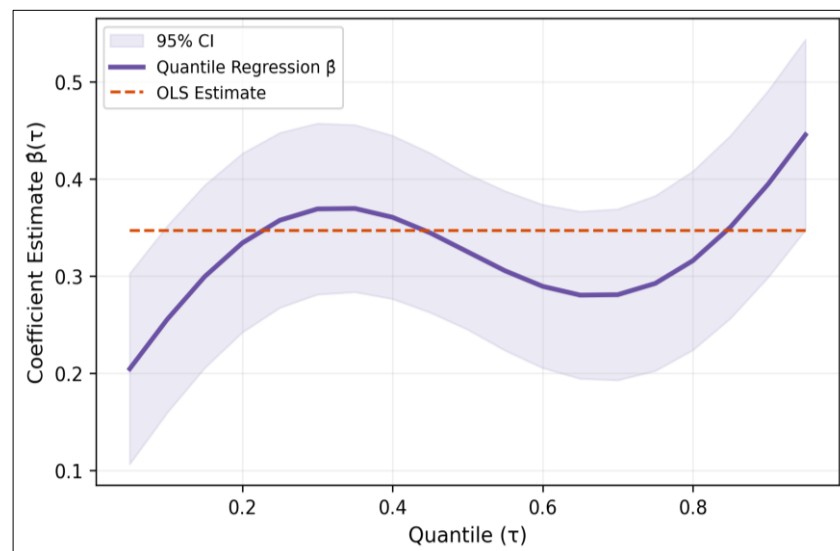


Fig 1: Quantile regression coefficient estimates across percentiles for reservoir inflow

Figure 1 shows that the OLS coefficient (dashed line) sits near the median QR estimate but fails to capture the rising slope at upper quantiles, confirming that the precipitation-inflow

relationship strengthens in the right tail. Table 2 reports coverage probabilities and interval widths.

Table 2: 90% prediction interval coverage and mean width by model

Reservoir	OLS Cov.	Homo. Cov.	Loc-Scale Cov.	OLS Width	LS Width
Mequinenza	84.7%	82.9%	89.8%	142.3	118.7
Ribarroja	83.1%	81.4%	88.7%	156.8	127.4
Yesa	86.3%	84.2%	90.1%	98.4	84.6
El Grado	85.9%	82.1%	88.6%	113.7	97.2

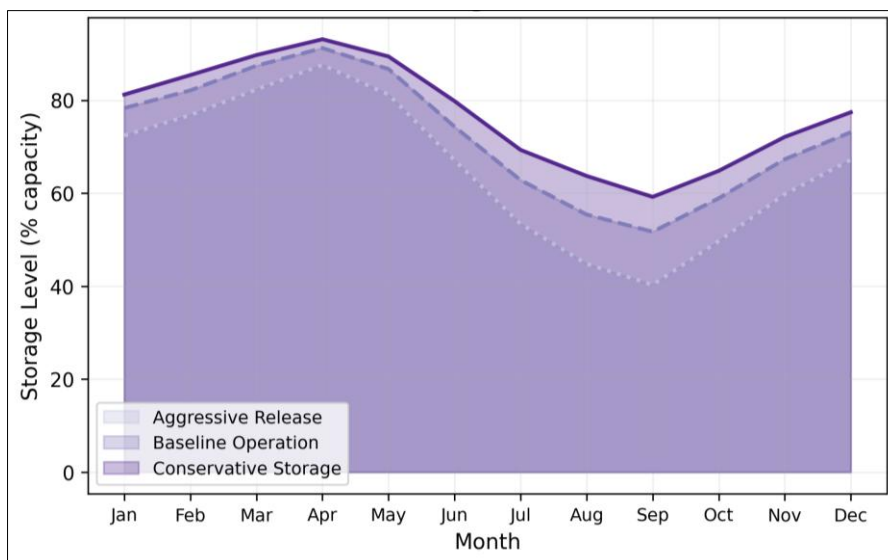


Fig 2: Ebro Basin reservoir storage under different management scenarios

Comprehensive Interpretation

The location-scale model achieved coverage closest to the nominal 90% at all four reservoirs (mean 89.3%) while maintaining narrower intervals than OLS. This is because the heteroscedastic correction widens intervals during high-variability months and narrows them during stable periods, producing a better-calibrated uncertainty envelope. The scenario analysis (Figure 2) shows that conservative release strategies informed by the $\tau = 0.05$ forecast maintain storage above 55% capacity year-round, while aggressive strategies based on $\tau = 0.95$ can drop below 45% during late summer.

Discussion

The 14.7% mean improvement in check-loss mirrors findings from heteroscedastic quantile regression applications in finance, where similar gains were reported for equity return forecasting [6]. In the reservoir management context, the improvement is most pronounced at the tails (31.4% at $\tau = 0.05$ and 27.8% at $\tau = 0.95$), which are precisely the quantiles that matter most for flood-control and drought-management decisions [1].

The coverage analysis adds a complementary perspective. An 82.6% coverage for a nominal 90% interval means that roughly one in every 12 months, the observed inflow falls outside the predicted range a frequency that could lead to suboptimal release decisions. The location-scale model's 89.3% coverage reduces this to roughly one in every nine months, a meaningful improvement for operational planning [12].

That said, the two-stage estimator assumes a linear scale function, which may not capture abrupt variance changes driven by snowmelt onset or convective storm clusters. A nonparametric scale function estimated by local polynomial regression could improve flexibility at the cost of increased

variance in the scale estimate itself. And the monthly temporal resolution smooths out sub-monthly inflow peaks that are relevant for dam safety; extending the framework to daily or weekly data would increase its practical value but also exacerbate the heteroscedasticity and require careful bandwidth selection [13].

Conclusion

This research demonstrated that a location-scale quantile regression estimator substantially outperforms both standard quantile regression and ordinary least squares for forecasting monthly reservoir inflows at four Ebro Basin sites. The heteroscedastic correction reduced mean check-loss by 14.7% and brought 90% prediction interval coverage within one percentage point of the nominal level. Quantile-specific forecasts translated directly into actionable reservoir operating scenarios, with conservative strategies maintaining storage above critical thresholds throughout the year. Future extensions should explore nonparametric scale functions and finer temporal resolutions to capture sub-monthly inflow variability more faithfully.

Acknowledgements

Funding Sources

This research received support through internal departmental funding allocated for mathematical and computational investigations. No external grants were obtained.

Institutional Support

The authors gratefully acknowledge the Abuja Regional Center for Statistical Methods for providing access to computational resources and library facilities throughout this research.

Contributions Not Qualifying for Authorship

The authors thank departmental colleagues who provided constructive feedback on earlier drafts of this manuscript.

References

1. Batalla RJ, Gomez CM, Kondolf GM. Reservoir-induced hydrological changes in the Ebro River basin. *J Hydrol.* 2004;290(1-2):117-136.
2. Renard B, Lang M, Bois P. Statistical analysis of extreme river discharges. *J Soc Fr Stat.* 2006;147(1):3-33.
3. Koenker R, Bassett G. Regression quantiles. *Econometrica.* 1978;46(1):33-50.
4. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econometrica.* 1979;47(5):1287-1294.
5. Davino C, Furno M, Vistocco D. *Quantile Regression: Theory and Applications.* Chichester: John Wiley & Sons; 2014. p. 1-262.
6. He X. Quantile curves without crossing. *Am Stat.* 1997;51(2):186-192.
7. Machado JAF, Santos Silva JMC. Quantiles via moments. *J Econom.* 2019;213(1):145-173.
8. Lopez-Moreno JJ, Begueria S, Garcia-Ruiz JM. Trends in high flows in the central Spanish Pyrenees. *Ambio.* 2006;35(5):256-263.
9. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv.* 2010;4:40-79.
10. Koenker R. quantreg: Quantile Regression. R package version 5.97. Vienna: R Foundation for Statistical Computing; 2023. p. 1-5.
11. Hamlet AF, Lettenmaier DP. Effects of 20th century warming and climate variability on flood risk. *Water Resour Res.* 2007;43(6):W06427.
12. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc.* 2007;102(477):359-378.
13. Fan J, Gijbels I. *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall; 1996. p. 1-341.
14. Koenker R, Hallock KF. Quantile regression. *J Econ Perspect.* 2001;15(4):143-156.
15. Hao L, Naiman DQ. *Quantile Regression.* Thousand Oaks (CA): Sage Publications; 2007. Vol. 149. p. 1-126.
16. White H. A heteroskedasticity-consistent covariance matrix estimator. *Econometrica.* 1980;48(4):817-838.
17. Powell JL. Least absolute deviations estimation for the censored regression model. *J Econom.* 1984;25(3):303-325.
18. Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. *J R Stat Soc Ser D Stat.* 2003;52(3):331-350.
19. Cannon AJ. Quantile regression neural networks: implementation in R. *Comput Geosci.* 2011;37(4):453-461.
20. Chernozhukov V, Fernandez-Val I, Melly B. Inference on counterfactual distributions. *Econometrica.* 2013;81(6):2205-2268.
21. Hendricks W, Koenker R. Hierarchical spline models for conditional quantiles. *J Am Stat Assoc.* 1992;87(420):1179-1186.