The Journal of Open Source Software

gkwreg: An R Package for Generalized Kumaraswamy Regression Models for Bounded Data

José Evandeilton Lopes [1] and **Wagner Hugo Bonat** [1]

1 Paraná Federal University, Brazil

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Mehmet Hakan Satman

Reviewers:

- @JonasMoss
- @faosorios

Submitted: 09 July 2025
Published: unpublished

Summary

gkwreg is an R package for fitting regression models to data restricted to the unit interval $(0, 1)$, such as proportions, rates, and indices. The package implements the flexible five-parameter Generalized Kumaraswamy (GKw) distribution and its seven main subfamilies, including the widely used Beta and Kumaraswamy distributions. A key feature of gkwreg is its use of the Template Model Builder (TMB) framework, which leverages automatic differentiation and C++ templates **to provide fast, stable, and accurate maximum likelihood estimation**. This overcomes the significant computational challenges typically associated with such complex multiparametric models, making them accessible for practical application. The package provides a user-friendly interface with standard R methods for model specification, inference, and diagnostics.

Statement of need

Statistical modeling of data bounded in the interval $(0, 1)$ is frequent across fields such as economics, epidemiology, and social sciences. Traditional methods like variable transformations followed by linear regression often present interpretability issues and fail near boundary points.

Direct modeling using distributions defined on $(0, 1)$ is preferable. While the Beta distribution is commonly used, it can be insufficient for complex patterns and lacks a closed-form cumulative distribution function (CDF). The Kumaraswamy (Kw) distribution (Kumaraswamy, 1980) offers an analytically simple CDF, yet its two-parameter form may be overly restrictive. To overcome these limitations, the Generalized Kumaraswamy (GKw) distribution, a flexible, five-parameter family incorporating the Beta and Kw distributions introduced by (Carrasco, Ferrari, & Cordeiro, 2010) was developed. However, practical applications of GKw in regression contexts have faced computational challenges. Its complex likelihood function makes Maximum Likelihood Estimation (MLE) computationally demanding and unstable, necessitating efficient and user-friendly computational tools.

The R package gkwreg (Lopes, 2025) addresses this need. Built on the Template Model Builder (TMB) package (Kristensen, Nielsen, Berg, Skaug, & Bell, 2016), it leverages automatic differentiation (AD) in C++ to efficiently compute gradients and Hessians, significantly enhancing speed, accuracy, and stability of MLE, especially when distribution parameters vary with covariates. gkwreg offers an intuitive interface aligned with standard R modeling conventions. Its integration with the multi-part formula syntax of the Formula package (Zeileis & Croissant, 2010) allows flexible specification of regression structures. Additionally, it provides comprehensive S3 methods (`summary()`, `predict()`, `plot()`, `residuals()`) and randomized quantile residuals (Dunn & Smyth, 1996) for model diagnostics, facilitating robust goodness-of-fit assessments. In complement, methods like `p*`, `d*`, `r*` and `q*` (eg. `dgkw()`) for seven GKw sub families were also implemented.

Mathematics

The Probability Density Function (PDF) of the five-parameter Generalized Kumaraswamy (GKw) distribution is given by:

$$f(y;\boldsymbol{\theta}) = \frac{\lambda\,\alpha\,\beta\,y^{\alpha-1}}{B(\gamma,\delta+1)}\,(1-y^{\alpha})^{\beta-1}\,\left[1-(1-y^{\alpha})^{\beta}\right]^{\gamma\lambda-1}\,\left\{1-\left[1-(1-y^{\alpha})^{\beta}\right]^{\lambda}\right\}^{\delta}$$

where $\boldsymbol{\theta} = (\alpha,\beta,\gamma,\delta,\lambda)^\top$ is the vector of positive shape parameters and $B(\cdot,\cdot)$ is the beta function.

Model diagnostics in `gkwreg` are primarily based on randomized quantile residuals, defined as:

$$r_i^Q = \Phi^{-1}(F(y_i;\hat{\boldsymbol{\theta}}_i))$$

where $F(y_i;\hat{\boldsymbol{\theta}}_i)$ is the fitted CDF evaluated at observation y_i with estimated parameters $\hat{\boldsymbol{\theta}}_i$, and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. If the model is correctly specified, these residuals should follow a standard normal distribution.

References

Carrasco, J. M. F., Ferrari, S. L. P., & Cordeiro, G. M. (2010). A new generalized Kumaraswamy distribution. *arXiv preprint arXiv:1004.0911*.

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, *5*(3), 236–244. doi:10.1080/10618600.1996.10474708

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, *70*(5), 1–21. doi:10.18637/jss.v070.i05

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, *46*(1-2), 79–88. doi:10.1016/0022-1694(80)90036-0

Lopes, J. (2025). *Gkwreg: An r package for generalized kumaraswamy regression models for bounded data*.

Zeileis, A., & Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, *34*(1), 1–13. doi:10.18637/jss.v034.i01