

Kumaraswamy regression modeling for Bounded Outcome Scores

Soudabeh Hamed-Shahraki¹, Aliakbar Rasekhi², Mohammad Reza Eshraghian³, Mir Saeed Yekaninejad^{3*}, Amir H Pakpour⁴



* Corresponding Author

1. Department of Epidemiology and Biostatistics, School of Public Health, Zabol University of Medical Sciences, Zabol, Iran, s.hsmedy1@gmail.com
2. Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran, rasekhi@modares.ac.ir
3. Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran, yekaninejad@yahoo.com
4. Social Determinants of Health Research Center, Qazvin University of Medical Sciences, Shahid Bahounar, Qazvin, Iran, pakpour_amir@yahoo.com

Abstract

In this paper, we use a regression model for modeling bounded outcome scores (BOS), where the outcome is Kumaraswamy distributed. Similar to the Beta distribution, this distribution can take a variety of shapes while being computationally easier to use. Thus, it is deemed as a suitable alternative distribution to the Beta in modeling bounded random processes. In the proposed model, the median of a bounded response is modeled by the linear predictors which are defined through regression parameters and explanatory variables. We obtained the maximum likelihood estimates (MLEs) of the parameters, provided closed-form expressions for the score functions and Fisher information matrix, and presented some diagnostic measures. We conducted Monte Carlo simulations to investigate the finite-sample performance of the MLEs of the parameters. Finally, two practical applications of this model to the real data sets are presented and discussed.

Keywords: Bounded outcome score; Kumaraswamy distribution; Beta regression; maximum likelihood estimation; diagnostic analysis.

1. Introduction

Bounded outcome scores (BOS) are measures restricted to a finite interval and are common in many health and medical disciplines. Visual Analogue Scale (VAS), which is often used to assess the severity of pain in pain relief studies (Wewers and Lowe, 1990); the PASI score that is used to record intensity psoriasis (Hu, Yeilding, Davis, and Zhou, 2011); the Barthel index for Activities of Daily Living (ADL) in stroke patients (Lesaffre, Rizopoulos, and Tsonaka, 2006); quality-of-life (QoL) measures in survey studies involving health assessment (Arostegui, Núñez - Antón, and Quintana, 2007; Hunger, Baumret, and Holle, 2011; Hunger, Döring, and Holle, 2012) are well-known instances of BOS measure outcomes.

The distributions of these outcomes can often take different shapes including unimodal, U-shaped, and J-shaped. Due to the variety of their distributions, BOS typically display heteroscedasticity, where the variance is smaller near the extremes, and asymmetry. Therefore, usual regression models such as simple linear or nonlinear regression models, are not suitable for these outcomes (Kieschnick and McCullough, 2003; Xu, Samtani, Yuan, and Nandy, 2014). Various strategies have been proposed for modeling BOS (Bottai, Cai, and McKeown, 2010; Hutton and Stanghellini, 2011; Kieschnick and McCullough, 2003; Lesaffre et al., 2006; Smithson and Verkuilen, 2006; Verkuilen and Smithson, 2012; Xu et al., 2014; Xu et al., 2013). Beta distribution is a well-known distribution which is widely used

to model double bounded data. The popularity of Beta distribution lies in the high flexibility of its density function which can take many shapes (e.g. unimodal, uniantimodal, increasing, decreasing, and uniform) through changing its parameter. Beta regression model, introduced by Ferrari and Cribari-Neto (Ferrari and Cribari-Neto, 2004) has been used in various disciplines to model random variables limited on a finite interval.

An alternative to the Beta distribution, which possesses similar characteristics, yet easier to use computationally is Kumaraswamy (“Kum for short”) distribution (Kumaraswamy, 1980). Kum distribution was proposed to model hydrological data, but is in fact useful for modeling many other outcomes that have lower and upper bounds such as heights of individuals, scores obtained in a test, atmospheric temperatures, and so on.

A random variable \tilde{Y} following a Kum distribution on the interval $[a, b]$ with respective shape parameters $p > 0$ and $q > 0$, is denoted by $Kum(p, q, a, b)$ and has a density function given by

$$f(\tilde{y}; p, q) = \frac{pq}{b-a} \left(\frac{\tilde{y}-a}{b-a} \right)^{p-1} \left[1 - \left(\frac{\tilde{y}-a}{b-a} \right)^p \right]^{q-1} ; \tilde{y} \in (a, b) \quad (1)$$

The mean and variance of this distribution are respectively give by:

$$E(\tilde{Y}) = a + (b-a)qB\left(1 + \frac{1}{p}, q\right) \quad (2)$$

and

$$\text{Var}(\tilde{Y}) = (b-a)^2 \left\{ qB\left(1 + \frac{2}{p}, q\right) - \left[qB\left(1 + \frac{1}{p}, q\right) \right]^2 \right\} \quad (3)$$

where $B(., .)$ is the Beta function.

The Kum distribution has many of the properties of the Beta distribution with some advantages in terms of tractability. Jones (Jones, 2009) specified the basic properties of the Kum distribution and discussed some similarities and differences between the Beta and Kum distributions. According to him, both distributions are very flexible and can take many different shapes depending on the values of their parameters. The author also emphasized several advantages of the Kum distribution over the Beta distribution: the normalizing constant is very simple; the distribution has a simple explicit formulae and quantile functions which do not involve any special functions; a simple formula for random variate generation; explicit formulae for moments of order statistics and L-moments. In addition, Jones noted that the Beta distribution has the following advantages over the Kum distribution: simpler formulae for moments and moment-generating function; a one-parameter sub-family of symmetric distributions; simpler moment estimation and more ways of generating the distribution via physical processes. Due to the similarity of Kum and Beta distributions and some advantages of Kum to the Beta distribution, including its explicit expression of the distribution and quantile functions, the Kum distribution has attracted some attention in the literature (Cordeiro, Nadarajah, and Ortega, 2012; Cordeiro, Ortega, and Nadarajah, 2010; de Pascoa, Ortega, and Cordeiro, 2011; Kızılaslan and Nadar, 2016). This study aimed to use the Kum distribution for modeling BOS. However, since in the standard parameterization of the Kum distribution, shape parameters are functions of the covariates, formulation of regression modeling and interpretation of the parameters are more difficult.

In addition, the lack of simple closed-form expressions for the mean and variance of the Kum distribution, given in (2) and (3), has limited its use for modeling purposes. On the other hand, the median of this distribution has the following simple expression,

$$\text{md}(\tilde{Y}) = \tilde{\omega} = a + (b-a) \left(1 - 0.5^{\frac{1}{q}} \right)^{\frac{1}{p}} \quad (4)$$

Therefore, we consider the median-based re-parameterization proposed by (Mitnik and Baek, 2013) aiming to facilitate its use in regression models. In order to develop the proposed model, we will follow the generalized linear models (McCullagh and Nelder, 1989) and Beta regression (Ferrari and Cribari-Neto, 2004) methodologies,; although as stated before, we will be modeling the median instead of the mean.

The paper unfolds as follows. Section 2 presents the Kum regression model and discusses the maximum likelihood estimation of this model. A Monte Carlo simulation study for assessing the finite sample performance of the conditional maximum likelihood approach is performed in Section 3. Finally, two applications using real data are presented in Section 4 and concluding remarks are given in Section 5.

2. Kum regression model and maximum likelihood estimator (MLE)

We use the general theory of maximum likelihood estimation. The estimates in this approach have properties such as asymptotic normality, consistency, and efficiency (see Ferguson, 2017; and Millar, 2011).

Let a random variable \tilde{Y} has the Kum distribution with density given by (1). With $Y = (\tilde{Y} - a)/(b - a)$, its median-based re-parameterization density function (Mitnik and Baek, 2013) is denoted by $Kum(p, \omega)$ which is given by

$$f(y) = \left(\frac{1}{b-a}\right) \frac{p \log(0.5)}{\log(1-\omega^p)} (y)^{p-1} (1-y^p)^{\frac{\ln(0.5)}{\ln(1-\omega^p)}-1} ; 0 < y < 1, \quad (5)$$

where $\omega = (\tilde{\omega} - a)/(b - a)$, $0 < \omega < 1$, and $p > 0$.

This form of density function allows us to use Kum distribution and develop models analogous to the Beta regression models. With a Kum-distributed dependent variable, $y_t \sim Kum(\omega_t, p)$, $t=1, \dots, n$; the model is obtained by assuming that the median of y_t can be written as:

$$g(\omega_t) = \sum_{i=1}^n x_{ti} \beta_i = \eta_t \quad (6)$$

where η_t is the linear predictor, $\beta = (\beta_1, \dots, \beta_k)^T$ is a vector of unknown regression parameters ($\beta \in \mathbb{R}^k$) and x_{t1}, \dots, x_{tk} are observations on k known covariates ($k < n$). Moreover, we assume that the link function $g(\cdot): (0,1) \rightarrow \mathbb{R}$ is a strictly monotonic and twice differentiable. Note that a number of various link functions such as logit, probit, log-log, and complementary log-log can be used. A comparison of these link functions can be found in Atkinson (1985, Ch. 7) and (Atkinson and Atkinson, 1985; McCullagh and Nelder, 1989). Our focus will be on the logit link, $g(\omega) = \log\left(\frac{\omega}{1-\omega}\right)$, which is one of the more popular link function that can be used. The popularity of the logit function is due to its straightforward interpretation, since the resulting regression coefficients can be interpreted as odds ratios.

Let $\mathbf{y} = (y_1, \dots, y_n)^T$ be a random sample of size n from Kum distribution and $\boldsymbol{\theta} = (\beta^T, p)^T$ be a $(k+1)$ -dimensional parameter vector. The log-likelihood function for the Kum regression model can be written as:

$$l(\boldsymbol{\theta}) = l(\beta, p) = \sum_{t=1}^n l_t(\omega_t, p), \quad (7)$$

where

$$l_t(\omega_t, p) = \log(p) - \log(b - a) + \log\left(\frac{\log 0.5}{\log(1-\omega_t^p)}\right) + (p-1) \log y_t + \left(\frac{\log 0.5}{\log(1-\omega_t^p)} - 1\right) \log(1 - y_t^p)$$

with $\omega_t = g^{-1}(\eta_t)$ as defined in (6) is a function of β . The MLE can be calculated by maximizing the log-likelihood function (7).

The components of the score vector are obtained by differentiating the log-likelihood function with respect to the unknown parameters β and p . The derivative of the log-likelihood function with respect to the i^{th} element of β , is given, for $i=1, \dots, k$, as

$$U_{\beta}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \beta_i} = \sum_{t=1}^n \frac{\partial l_t(\omega_t, p)}{\partial \omega_t} \frac{d\omega_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i}$$

where $\frac{d\omega_t}{d\eta_t} = \frac{1}{g'(\omega_t)}$ and $\frac{\partial \eta_t}{\partial \beta_i} = x_{ti}$, and

$$\frac{\partial l_t(\omega_t, p)}{\partial \omega_t} = \left(\frac{p\omega_t^{p-1}}{(1-\omega_t^p) \log(1-\omega_t^p)}\right) \left[1 + \frac{\log 0.5}{\log(1-\omega_t^p)} \log(1 - y_t^p)\right] \quad (8)$$

thus

$$U_{\beta}(\boldsymbol{\theta}) = \sum_{t=1}^n \left\{ \left(\frac{p\omega_t^{p-1}}{(1-\omega_t^p) \log(1-\omega_t^p)}\right) \left[1 + \frac{\log 0.5}{\log(1-\omega_t^p)} \log(1 - y_t^p)\right] \right\} \frac{1}{g'(\omega_t)} x_{ti}. \quad (9)$$

By

$$q_t = \frac{\log 0.5}{\log(1-\omega_t^p)}, \quad \text{and} \quad c_t = \left\{ \left(\frac{p\omega_t^{p-1}}{(1-\omega_t^p) \log(1-\omega_t^p)}\right) [1 + q_t \log(1 - y_t^p)] \right\}, \quad (10)$$

The matrix form of (9) is given by

$$U_{\beta}(\boldsymbol{\theta}) = X^T T c, \quad (11)$$

with X being an $n \times k$ matrix whose t^{th} entry is x_t^T , $T = \text{diag}\{1/g'(\omega_1), \dots, 1/g'(\omega_n)\}$ and $c = (c_1, \dots, c_n)^T$ with equation (10).

Similarly, the score function for p can be obtained by differentiating the log-likelihood function with respect to p , yielding

$$U_p(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial p} = \sum_{t=1}^n \frac{\partial l_t(\omega_t, p)}{\partial p}$$

where

$$\frac{\partial l_t(\omega_t, p)}{\partial p} = \frac{1}{p} + \log y_t + \left(\frac{\omega_t^p \log \omega_t}{(1-\omega_t^p) \log(1-\omega_t^p)} \right) \left[1 + \frac{\log 0.5}{\log(1-\omega_t^p)} \log(1 - y_t^p) \right] - \left(\frac{\log 0.5}{\log(1-\omega_t^p)} - 1 \right) \frac{y_t^p \log y_t}{1-y_t^p}$$

and by using (10), it can be written as follows:

$$U_p(\boldsymbol{\theta}) = \frac{n}{p} + \sum_{t=1}^n \left\{ \log y_t + c_t \frac{\omega_t \log \omega_t}{p} - (q_t - 1) \frac{y_t^p \log y_t}{1-y_t^p} \right\}. \quad (12)$$

The MLEs of β and p are obtained as the solution of the equations $U_\beta(\boldsymbol{\theta}) = 0$ and $U_p(\boldsymbol{\theta}) = 0$. These equations do not have a closed-form and can be obtained through standard numerical optimization procedures, e.g. the Newton or quasi-Newton algorithm such as Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Nocedal and Wright, 1999; Press, Teukolsky, Vetterling, and Flannery, 2007). An iterative optimization algorithm requires initial value. The starting value of the regressors parameter β was selected from the maximum likelihood estimate from a Beta regression.

In the next step, we obtain an expression for Fisher information matrix as minus the expected value of the second derivatives of the log-likelihood. It is shown in Appendix A that Fisher information matrix is given by:

$$K(\boldsymbol{\theta}) = \begin{pmatrix} K_{\beta\beta} & K_{\beta p} \\ K_{p\beta} & K_{pp} \end{pmatrix}, \quad (13)$$

where $K_{\beta\beta} = X^T R^2 T^2 X$, $K_{\beta p} = K_{p\beta}^T = -X^T B T \mathbf{1}$, $K_{pp} = -tr(D)$, with

$R = \text{diag}\{r_1, \dots, r_n\}$, with r_t in equation (A1),

$B = \text{diag}\{b_1, \dots, b_n\}$, with equation (A3),

$D = \text{diag}\left\{ \frac{\partial^2 l_1(\omega_1, p)}{\partial p^2}, \dots, \frac{\partial^2 l_n(\omega_n, p)}{\partial p^2} \right\}$,

$\mathbf{1}$ is an $n \times 1$ vector of ones

and $tr(\cdot)$ is the trace function.

Under the usual regularity conditions and provided that the sample size is large enough, it can be shown that the asymptotic distribution of the MLE $\hat{\boldsymbol{\theta}} = (\hat{\beta}^T, \hat{p})$ is the multivariate normal distribution with mean $\boldsymbol{\theta} = (\beta^T, p)$ and variance-covariance matrix $K^{-1}(\boldsymbol{\theta})$, that is,

$$\hat{\boldsymbol{\theta}} \sim N_{k+1}(\boldsymbol{\theta}, K^{-1}(\boldsymbol{\theta}))$$

where $K^{-1}(\boldsymbol{\theta})$ is the expected information matrix which can be computed numerically.

In large sample inference in the Kum regression model, the $100(1 - \alpha)\%$ asymptotic confidence interval for β_i , $i = 1, \dots, k$, is given by $\hat{\beta}_i \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) se(\hat{\beta}_i)$, where $se(\hat{\beta}_i)$ is the asymptotic standard error of the maximum likelihood estimator of $\hat{\beta}_i$ obtained from the inverse of Fishers information matrix evaluated at the maximum likelihood estimates.

3. Numerical results

In this section, we present the results of Monte Carlo simulation, where we study the finite-sample distributions of the MLEs of the parameters β and p along with the Kum regression model in which

$$g(\omega_t) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}, \quad t = 1, \dots, n,$$

where $g(\cdot)$ is the logit link function.

The covariate values are obtained as random draws from the following distributions: covariate x_{1t} is generated from a continuous Uniform $U(1, 5)$ distribution, while x_{2t} is generated from a Binomial distribution with three number of trials and 0.5 success probability in each trial ($x_{2t} \sim \text{Bin}(3, 0.5)$). The covariate values remain constant throughout the simulations. We take $\beta_0 = -6.0$, $\beta_1 = 1.0$ and $\beta_2 = 2.0$, which resulted in median values between zero and one. The value of p is also set at $p = 2$.

We take samples of sizes $n = 50, 100, 200$, and 500 . For each sample size, we perform 1000 simulations and compute the mean of the estimates, percentage relative bias and mean square errors. It should be noted that the percentage relative bias is defined as the ratio between the bias and the true parameter value times 100. Table 1 presents the simulation results, and the plot of corresponding percentage relative bias (RB %) for the different sample sizes n , is

shown in Figure 1. We note that the mean of estimates is close to the true parameter values, the percentage relative bias and mean square errors decrease as sample size increases. Overall, we observe that the estimates improve with increasing sample size.

Table 1 Estimation of the mean parameters, percentage relative bias (RB%) and mean square errors (MSE) for Kum regression model (N=1000)

Parameter	$\beta_0 = -6$	$\beta_1 = 1$	$\beta_2 = 2$	$p = 2$
n=50				
Mean	-6.0337	1.0035	2.0229	2.1503
RB (%)	0.5619	0.3473	1.1474	7.5161
MSE	0.1306	0.0123	0.0393	0.1336
n=100				
Mean	-6.0195	1.0038	2.0084	2.0645
RB (%)	0.3252	0.3755	0.4201	3.2268
MSE	0.1111	0.0084	0.0152	0.0539
n=200				
Mean	-6.0084	1.0006	2.0073	2.0364
RB (%)	0.1396	0.0583	0.3667	1.8219
MSE	0.0330	0.0030	0.0088	0.0234
n=500				
Mean	-5.9990	0.9997	2.0008	2.0180
RB (%)	-0.0173	-0.0303	0.0415	0.8991
MSE	0.0143	0.0013	0.0023	0.0087

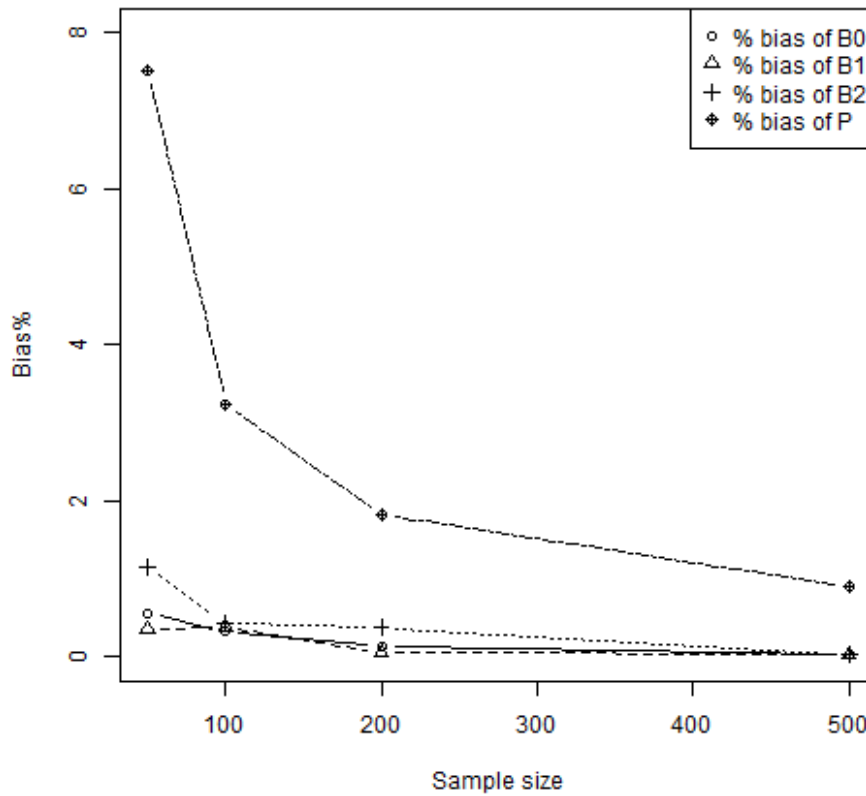


Figure 1: Percentage relative bias (RB%) versus sample size of $\beta_0, \beta_1, \beta_2, p$.

4. Application

In this Section, we present two applications of the Kum regression model described in Section 2.

4.1. Health-related quality of life in patients with epilepsy

The first application is based on an empirical dataset that comes from a longitudinal study on the quality of life in patients with epilepsy in Iran from March 2014 to December 2015.

Epilepsy is one of the most common chronic brain disorders that has a considerable impact on person's QoL. We will examine the relationship between medication adherence and seizure severity on QoL. Medication Adherence and seizure severity were assessed by the Medication Adherence Report Scale (MARS-5) and the Liverpool Seizure Severity Scale (LSSS) questionnaires, respectively. The quality of life in epilepsy (QOLIE-31) was used for measuring health-related QoL in epileptic patients. QOLIE-31 has seven sub-scales, which are seizure worry, cognitive functioning, energy/fatigue, emotional well-being, social functioning, medication effects, and overall QoL (Cramer et al., 1998). The subscales are constrained to a bounded interval (e.g. at 0 and 1) and a higher score indicates better QoL. Our interest is to model one of the sub-scales of QoL, namely social functioning, as a function of medication adherence and seizure severity. Table 2 lists some summary statistics for the social functioning. We note that the median of social functioning is greater than its mean, indicating left skewness of its distribution.

Table 2 Statistical description of social functioning

Variable	Min.	Max.	Mean	Med.	SD	Skewness	Kurtosis
Social Function	.001	.999	.794	.874	.202	-0.718	-0.244

Given that the outcome, social functioning, is bounded between 0 and 1 and its distribution is negatively skewed, the traditional regression models may not be suitable for regression modeling. Therefore, the Beta and Kum regression models which is proposed in Section 2 will be used instead. We determine the MLEs of the model parameters using the quasi-Newton optimization algorithm known as BFGS with analytical derivatives. The choice of starting values for the unknown parameters is based on the suggestion made in the second section. The MLEs of the model parameters, their estimated standard errors, and the Akaike Information Criterion (AIC) are presented in Table 3. We note that the AIC value for the Kum model is smaller than the AIC value for Beta model. So, based on the AIC criterion, the Kum distribution is preferable in fitting these data than the Beta distribution.

Table 3 Parameter estimates with logit link

	Beta regression			Kum regression		
	Estimate	Std. error	<i>p</i>	Estimate	Std. error	<i>p</i>
Intercept	1.516*	0.175	<0.001	2.464*	0.345	<0.001
LSSS score	-0.007*	0.002	<0.001	-0.014*	0.004	<0.001
MARS-5 score	0.028*	0.007	<0.001	0.059*	0.015	<0.001
Phi	1.929*	0.105	<0.001	1.718*	0.116	<0.001
AIC	-1815.36			-1817.39		

LSSS, Liverpool Seizure Severity Scale; MARS-5, Medication Adherence Report Scale; AIC, Akaike information criterion

4.2. Food expenditure

The next application is based on the food expenditure data analyzed by (Ferrari and Cribari-Neto, 2004), the source of the data is (Griffiths, Carter Hill, and Judge, 1993). We model the proportion of income spent on food as a function

of the level of income and the number of persons in the household using Beta and Kum regression models. The link function used is the logit link.

Estimates of the parameters, their standard errors, and the Akaike Information Criterion (AIC) are given in Table 4. The values in Table 4 show that the estimated coefficients associated with the covariates are quite similar in both models. The lower AIC value for the Kum model indicates that this model is a better fit for the current data than the Beta model.

Table 4 Parameter estimates with logit link

	Beta regression			Kum regression		
	Estimate	Std. error	<i>p</i>	Estimate	Std. error	<i>p</i>
Intercept	-0.622*	0.224	0.005	-0.844*	0.199	<0.0001
Total family income	-0.012*	0.003	<0.0001	-0.009*	0.002	<0.0001
Number of residents	0.118*	0.035	0.001	0.135*	0.028	<0.0001
Phi	35.733*	8.080	<0.0001	4.949*	0.661	<0.0001
AIC		-82.667			-89.717	

5. Concluding remarks

In this paper we introduced a regression model for the double bounded dependent outcomes. The underlying assumption was that the response variable follows a Kum distribution. In the proposed regression model, the median of the dependent variable was modeled by a linear predictor that was defined by regression parameters and explanatory variables. Inference about the Kum regression model was discussed and the parameter estimation was performed via maximum likelihood. The closed-form expressions were obtained for the score functions and the Fisher's information matrix. Based on the asymptotic normality of the maximum likelihood estimator, confidence interval and hypothesis testing were obtained. Furthermore, some diagnostic measures were provided. A Monte-Carlo simulation study was performed to evaluate the finite-sample performance of the MLEs of β and p with the Kum regression model. The simulation results showed that the MLE performs well even for small sample sizes. Finally, to illustrate its usefulness, two applications of the Kum regression model to the empirical real data were presented, where one of them was based on a previously published results. We hope the current model could attract applications in several areas.

Acknowledgments:

The authors would like to thank the Research Vice-Chancellor of Tehran University of Medical Sciences for financial and support. This work was approved by the ethics committee of Tehran University of Medical Sciences (Approval No. IR.TUMS.SPH.REC.1396.2907).

References

1. Arostegui, I., Núñez-Antón, V., and Quintana, J. M. (2007). Analysis of the short form-36 (SF-36): the beta-binomial distribution approach. *Statistics in medicine*, 26(6), 1318-1342.
2. Atkinson, A. C. (1985). *Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis* (New York: Wiley).
3. Bottai, M., Cai, B., and McKeown, R. E. (2010). Logistic quantile regression for bounded outcomes. *Statistics in medicine*, 29(2), 309-317.
4. Cordeiro, G. M., Nadarajah, S., and Ortega, E. M. (2012). The Kumaraswamy Gumbel distribution. *Statistical Methods and Applications*, 21(2), 139-168.
5. Cordeiro, G. M., Ortega, E. M., and Nadarajah, S. (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute*, 347(8), 1399-1429.
6. Cramer, J. A., Perrine, K., Devinsky, O., Bryant-Comstock, L., Meador, K., and Hermann, B. (1998). Development and cross-cultural translations of a 31-item quality of life in epilepsy inventory. *Epilepsia*, 39(1), 81-88.
7. De Pascoa, M. A., Ortega, E. M., and Cordeiro, G. M. (2011). The Kumaraswamy generalized gamma distribution with application in survival analysis. *Statistical Methodology*, 8(5), 411-433.

8. Ferguson, T. S. (2017). *A Course in Large Sample Theory*. New York: Routledge, <https://doi.org/10.1201/9781315136288>.
9. Ferrari, S., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.
10. Griffiths, W. E., Hill, R. C., and Judge, G. G. (1993). *Learning and practicing econometrics*. John Wiley and Sons. Inc. New York, USA.
11. Hu, C., Yeilding, N., Davis, H. M., and Zhou, H. (2011). Bounded outcome score modeling: application to treating psoriasis with ustekinumab. *Journal of pharmacokinetics and pharmacodynamics*, 38(4), 497-517.
12. Hunger, M., Baumert, J., and Holle, R. (2011). Analysis of SF-6D index data: is beta regression appropriate? *Value in Health*, 14(5), 759-767.
13. Hunger, M., Döring, A., and Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC medical research methodology*, 12(1), 144.
14. Hutton, J. L., and Stanghellini, E. (2011). Modelling bounded health scores with censored skew-normal distributions. *Statistics in medicine*, 30(4), 368-376.
15. Jones, M. C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, 6(1), 70-81.
16. Kieschnick, R., and McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. *Statistical modelling*, 3(3), 193-213.
17. Kızılaslan, F., and Nadar, M. (2016). Estimation and prediction of the Kumaraswamy distribution based on record values and inter-record times. *Journal of Statistical Computation and Simulation*, 86(12), 2471-2493.
18. Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1-2), 79-88.
19. Lesaffre, E., Rizopoulos, D., and Tsonaka, R. (2007). The logistic transform for bounded outcome scores. *Biostatistics*, 8(1), 72-85.
20. McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, no. 37 in *Monograph on Statistics and Applied Probability*: Chapman and Hall.
21. Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB* (Vol. 111). John Wiley and Sons
22. Mitnik, P. A., and Baek, S. (2013). The Kumaraswamy distribution: median-dispersion re-parameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54(1), 177-192.
23. Nocedal, J., and Wright, S. (1999). *Numerical Optimization* Springer-Verlag. New York.
24. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press.
25. Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, 11(1), 54-71.
26. Verkuilen, J., and Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, 37(1), 82-113.
27. Wewers, M. E., and Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in nursing and health*, 13(4), 227-236.
28. Xu, X. S., Samtani, M., Yuan, M., and Nandy, P. (2014). Modeling of bounded outcome scores with data on the boundaries: Application to disability assessment for dementia scores in Alzheimer's disease. *The AAPS journal*, 16(6), 1271-1281. doi:10.1208/s12248-014-9655-y
29. Xu, X. S., Samtani, M. N., Dunne, A., Nandy, P., Vermeulen, A., De Ridder, F., and Initiative, A. s. D. N. (2013). Mixed-effects beta regression for modeling continuous bounded outcome scores using NONMEM when data are not on the boundaries. *Journal of pharmacokinetics and pharmacodynamics*, 40(4), 537-544.

Appendix A

In this appendix we provide observed information matrix for (β, p) in the class of kum regression models. The information matrix can be obtained as minus the expected value of the second derivatives of the log-likelihood.

Observed information matrix

The second-order partial derivative of equation (7) with respect to β s is given by

$$\begin{aligned} \frac{\partial^2 l(\beta, p)}{\partial \beta_i \partial \beta_j} &= \sum_{t=1}^n \frac{\partial}{\partial \omega_t} \left(\frac{\partial l_t(\omega_t, p)}{\partial \omega_t} \frac{d\omega_t}{d\eta_t} \right) \frac{d\omega_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_j} x_{ti} \\ &= \sum_{t=1}^n \left(\frac{\partial^2 l_t(\omega_t, p)}{\partial \omega_t^2} \frac{d\omega_t}{d\eta_t} + \frac{\partial l_t(\omega_t, p)}{\partial \omega_t} \frac{\partial}{\partial \omega_t} \frac{d\omega_t}{d\eta_t} \right) \frac{d\omega_t}{d\eta_t} x_{ti} x_{tj} \end{aligned}$$

where $\frac{d\omega_t}{d\eta_t} = \frac{1}{g'(\omega_t)}$ and $\frac{\partial \eta_t}{\partial \beta_j} = x_{tj}$.

$$\frac{\partial^2 l(\beta, p)}{\partial \beta_i \partial \beta_j} = \sum_{t=1}^n \left(\frac{\partial^2 l_t(\omega_t, p)}{\partial \omega_t^2} \right) \left(\frac{1}{g'(\omega_t)} \right)^2 x_{ti} x_{tj}$$

Now, by differentiating the equation (8) with respect to ω_t , and taking the expected value of the second derivative, since $E(\log(1 - y_t^p)) = -\frac{1}{q_t}$, yielding

$$E \left(\frac{\partial^2 l_t(\omega_t, p)}{\partial \omega_t^2} \right) = - \left(\frac{p \omega_t^{p-1}}{(1 - \omega_t^p) \log(1 - \omega_t^p)} \right)^2$$

Therefore

$$E \left(\frac{\partial^2 l(\beta, p)}{\partial \beta_i \partial \beta_j} \right) = - \sum_{t=1}^n \left[\left(\frac{p \omega_t^{p-1}}{(1 - \omega_t^p) \log(1 - \omega_t^p)} \right)^2 \left(\frac{1}{g'(\omega_t)} \right)^2 x_{ti} x_{tj} \right].$$

In matrix notation, we have

$$E \left(\frac{\partial^2 l(\beta, p)}{\partial \beta_i \partial \beta_j} \right) = -X^T R^2 T^2 X$$

with X being an $n \times k$ matrix whose t th row is x_t^T , $T = \text{diag}\{1/g'(\omega_1), \dots, 1/g'(\omega_n)\}$ and $R = \text{diag}\{r_1, \dots, r_n\}$, with

$$r_t = p m_t \tag{A1}$$

where $m = (m_1, \dots, m_n)^T$ with equation $m_t = \frac{\omega_t^{p-1}}{(1 - \omega_t^p) \log(1 - \omega_t^p)}$.

Analogously, the second derivative of the log-likelihood function with respect to β_i and p is given by

$$\frac{\partial^2 l_t(\beta, p)}{\partial \beta_i \partial p} = \sum_{t=1}^n \frac{\partial}{\partial \omega_t} \left(\frac{\partial l_t(\omega_t, p)}{\partial p} \right) \frac{d\omega_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i} = \sum_{t=1}^n \frac{\partial^2 l_t(\omega_t, p)}{\partial \omega_t \partial p} \frac{d\omega_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_i}$$

where

$$\begin{aligned} \frac{\partial^2 l_t(\omega_t, p)}{\partial \omega_t \partial p} &= \frac{[p \omega_t^{p-1} \log \omega_t + \omega_t^{p-1}](1 - \omega_t^p) \log(1 - \omega_t^p) - \omega_t^p \log \omega_t [-p \omega_t^{p-1} \log(1 - \omega_t^p) - p \omega_t^{p-1}]}{[(1 - \omega_t^p) \log(1 - \omega_t^p)]^2} [1 + q_t \log(1 - y_t^p)] + \\ &\left(\frac{\omega_t^p \log \omega_t}{(1 - \omega_t^p) \log(1 - \omega_t^p)} \right) \left(q_t \frac{p \omega_t^{p-1}}{(1 - \omega_t^p) \log(1 - \omega_t^p)} \log(1 - y_t^p) \right) - \frac{y_t^p \log y_t}{1 - y_t^p} \left(p q_t \frac{\omega_t^{p-1}}{(1 - \omega_t^p) \log(1 - \omega_t^p)} \right) \end{aligned} \tag{A2}$$

taking conditional expectation in (A2) and since $E(\log(1 - Y_t^p)) = -\frac{1}{q_t}$, $E \left(\frac{Y_t^p \log Y_t}{1 - Y_t^p} \right) = \frac{1 - \psi(q_t + 1) - \gamma}{p(q_t - 1)}$, with $\psi(\cdot)$

denotes the digamma function defined as $\psi(z) = \frac{d}{dz} \log(\Gamma(z))$, and γ is the Euler-Mascheroni constant ($\gamma = 0.57721566 \dots$), it follows that

$$E \left(\frac{\partial^2 l_t(\beta, p)}{\partial \beta_i \partial p} \right) = \sum_{t=1}^n b_t \frac{1}{g'(\omega_t)} x_{ti}$$

where

$$b_t = -p \omega_t \log \omega_t m_t^2 - p q_t m_t \left(\frac{1 - \psi(q_t + 1) - \gamma}{p(q_t - 1)} \right) \tag{A3}$$

In matrix form, we have

$$U_{\beta p}(\beta, p) = E \left(\frac{\partial^2 l_t(\beta, p)}{\partial \beta_i \partial p} \right) = X^T B T \mathbf{1}$$

with X being an $n \times k$ matrix whose t th is x_t^T , $B = \text{diag}\{b_1, \dots, b_n\}$ with equation (A3), $T = \text{diag}\{1/g'(\omega_1), \dots, 1/g'(\omega_n)\}$ and $\mathbf{1}$ is an $n \times 1$ vector of ones.

Finally, the second derivative of $l(\beta, p)$ with respect to p comes by differentiating the expression in equation (12) with respect to p . Therefore $E \left(\frac{\partial^2 l_t(\omega_t, p)}{\partial p^2} \right) = \sum_{t=1}^n d_t$, which, in matrix notation, can be written as

$$E \left(\frac{\partial^2 l_t(\omega_t, p)}{\partial p^2} \right) = \text{tr}(D)$$

where $D = \text{diag}\left\{\frac{\partial^2 l_1(\omega_1, p)}{\partial p^2}, \dots, \frac{\partial^2 l_n(\omega_n, p)}{\partial p^2}\right\}$ and $\text{tr}(\cdot)$ is the trace function.

Now, the Fisher information matrix for (β, p) as given in (13), can be obtained.