Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/140420

Support Vector Regression and Beta Distribution for Modeling Incumbent Party for Presidential Elections

C. R. Kikawa[1,2,*], *M. N. Ngungu*[3], *D. Ntirampeba*[2] *and A. Ssematimba*[4]

[1]Department of Mathematics and Statistics, Faculty of Health and Applied Sciences, Namibia University of Science and Technology, Brahms and Haydn Street, Private Bag 13388, Windhoek, Namibia
[2]Department of Economics and Statistics, Faculty of Economics and Management Sciences, Kabale University, P.o.Box 317, Kabale, Uganda
[3]Democracy, Governance and Service Delivery, Human Sciences Research Council, Private Bag 41, Pretoria 0001, South Africa
[4]Department of Mathematics, Faculty of Science, Gulu University, P.o.Box 166, Gulu, Uganda

Received: 18 Nov. 2019, Revised: 22 Feb. 2020, Accepted: 22 Apr. 2020
Published online: 1 Jul. 2020

Abstract: The present paper aims to model, predict, and explain presidential election results using selected quarterly macroeconomic indicators, i.e., gross national product, consumer price index, unemployment rate and gross national product from 1994-2017. We also seek to provide predictions of presidential winner prior to the elections based on the beta distribution and the support vector regression (SVR) as prediction models.Two models are primarily built based on beta distribution and SVR. Due to the forecasting aspect, model performance focuses on one goodness-of-fit measure, i.e., the prediction error rather than the squared correlation coefficient R^2 as it makes little sense in a practical regression perspective. The best model is the one with the least mean square error (*MSE*). In this effect it turns out that the SVR with kernel type encapsulated postscript eps radial has a mean square error of 0.006 on the test set and is a better model compared to the beta distribution model with a mean square error of 1.216. Thu, an accurate solution to prediction of presidential vote elections via SVR analysis is proposed.

Keywords: Proportions,Time series data, Predictions, Regression, Support Vector Machine

1 Introduction

Within the scientific framework in line with politics, the primary focus is to provide explanations concerning elections. Nevertheless, in some phenomena like presidential elections, the aim of predicting is important, as well. It is usually the case for the citizenry both academic and non-academic to ask themselves who the next president will be. These interlocutors tend to require predictions, not just explanations. They primarily want an accurate prediction for the winner of the forthcoming elections. In this work, we attempt to meet their quest. We consider two models that can be employed in modeling proportions, i.e. the beta distribution [1] and the support vector regression [2]. Their performance is compared using the goodness-of-fit measures as the measure square error (MSE) and the squared correlation coefficient (R^2) to ascertain which one gives more accurate predictions based on the selected macroeconomic time series data, i.e. the real gross domestic product (GDP), consumer price index (CPI), unemployment rate and the gross national product (GNP) from the post apartheid era, (1994-2017) in South Africa. Accordingly, we recommend an appropriate modeling framework that provides presidential election outcomes accurately.
The prsent paper aims to present true forecasting using predictor models. That is, to predict the winner before the elections take place. In that respect two prediction models are adopted, i.e. the beta distribution and the support vector regression (SVR). Due to our interest in prediction, particular emphasis is put on the goodness-of-fit measures, like the squared correlation coefficient, R^2, and prediction error given by the mean square error, [3,4]

* Corresponding author e-mail: crkikawa@kab.ac.ug

2 Modeling proportions

It is a hard task to select an appropriate regression model when the dependent variable is a proportion or percentage. When using ordinary linear regression, a big problem arises because the model can predict values that are not possible values less than 0 or greater than 1 particularly in voting scenarios. The other issue is that the relationship is nonlinear, because it turns out to be sigmoidal [5,6]. Several researchers envisage application of a linear regression as it is much simpler; however, this approach can be reasonable only in a few circumstances [7]. For a thorough understanding of the situations under which the linear regression could be justified, see [6]. There are other proportions data (i.e. data of discrete counts) derived from discrete counts of "successes" and "failures" in which "successes" are divided by the total counts [8]. Each observation is a percentage from 0 to 100%, or a proportion from 0 to 1. This type of data can be analyzed with beta regression or logistic regression.
For data that are inherently proportional (proportion data), one can not count "successes" or "failures", but just divide one continuous variable by a provided denominator value. Another instance of such proportion data is when a proportion is estimated by intuitive measurement [9]. This type of data can be analyzed with beta regression.

3 Beta Regression

Continuous bounded regressands, such as proportions and rates, exist in several areas of applied statistics. Linear regression prior to a logistic regression is commonly employed to explore such type of data. However, regression parameters from that type of regression are indirectly interpretable on the original response scale, as a consequence of Jensen's inequality [10]. Another approach to linear modeling prior to the logistic transformation is involved in a direct assessment of the bounded responses on their original scale. In this respect, the beta regression model has recently gained much appreciation because of the flexibility of the beta distribution in encompassing a number of distributional shapes over the unit interval [12,11,1]. The two-parameter probability density function of the beta distribution with shape parameters α and β, see Def[1].

Definition 1*A random variable X is said to follow the* ***beta distribution*** *with parameters* (α,β) *for some* $\alpha > 0$ *and* $\beta > 0$ *if*

$$Range(X) = (0,1)$$

, and for $x \in (0,1)$ *then*

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

. We write this as $X \sim Beta(\alpha,\beta)$.

In our study we have assumed that the vote proportions (percentages) of the incumbent party, the African National Congress (ANC) follow a beta distribution; $vote_{ANC} \sim Beta(\alpha,\beta)$.
The parameters α and β are symmetrically related by $f(x \mid \alpha,\beta) = f(1-x \mid \beta,\alpha)$. That is, when X has a beta distribution with parameters α and β, $1-X$ has a beta distribution with parameters β and α [14,15].

3.1 Estimation of the beta parameters

The beta distribution takes on various shapes and can be described by two shape parameters, α and β, which are usually cumbersome to estimate. However, the maximum likelihood and method of moments estimation approaches can be employed for the estimation problem. Nevertheless, the method of moments is much more straightforward [13,16]. The shape of the beta distribution can readily change with variation in the parameters [13, 15]. For a detailed pieces of literature on the estimation of beta parameters the interested reader should read [13,16].

4 Support Vector Machine-Regression

The application of support vector machine regression (SVM-R) in this study is motivated by a growing popularity of support vector machines (SVM) learning algorithms for regression problems [17,18]. SVMs are considered the best supervised learning algorithms by scholars and researchers of the 21st century in the data science arena.
The practical accomplishment of SVMs can be accredited to the core theoretical basis of Vapnik–Chervonenkis (VC)-theory [19,20] because SVM generalization performance does not depend on the dimensionality of the input space. Nevertheless, a number of SVM regression application practices are carried out by 'authority' users. Given that the quality of SVM models is based on a proper setting of SVM hyper-parameters, the core concern for users attempting to employ SVM regression is how to set these parameter values (to ensure good generalization performance) for a given data set [21]. In this study, the prediction of incumbent vote proportion problem focuses on the application of SVR, so we can use SVR for handling continuous values in the range $[0,1]$ instead of SVM.
Given that the ultimate goal of the support vector regression is to find a regression function, we then introduce $\varepsilon-$support vector regression, or the $\varepsilon-$insensitive loss function which we shall employ to forecast the incumbent presidential vote.

4.1 $\varepsilon-$Support Vector Regression

Given the training set

$$T=\{(x_1,y_1),...,(x_m,y_m)\}\in(R^n\times\gamma)^m, \quad (1)$$

where $x_i\in R^n$, $y_i\in\gamma=R$, $i=1,\dots,m$,

Linear problem

The application of linear $\varepsilon-$ support vector regression enhances establishing the problem as a convex quadratic programming problem (QPP), where coefficients ω and b are estimated by minimizing

$$\begin{aligned} \underset{\omega,b,\xi^{(*)}}{\text{minimize}} \quad & \frac{1}{2}\|\omega\|^2+C\Sigma_{i=1}^{m}(\xi_i+\xi_i^*), \\ \text{subject to} \quad & (\omega\cdot x_i)+b-y_i\le\varepsilon+\xi_i,\ i=1,\dots,m, \\ & y_i-(\omega\cdot x_i)-b\le\varepsilon+\xi_i^*,\ i=1,\dots,m, \\ & \xi_i^{(*)}\ge 0,\ i=1,\dots,m, \end{aligned} \quad (2)$$

where $(*)$ symbolizes both the vector with and without asterisks. $\xi^{(*)}=(\xi_1,\xi_1^*,\dots,\xi_m,\xi_m^*)^T$ is a slack variable and $C>0$ is a penalty parameter.

Now, we introduce the Lagrangian function to formulate the dual problem of the primal problem Equation (2),

$$\begin{aligned} L=&\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{m}(\xi_i+\xi_i^*)-\sum_{i=1}^{m}(\lambda_i\xi_i+\lambda_i^*\xi_i^*) \\ &-\sum_{i=1}^{m}\alpha_i(\varepsilon+\xi_i+y_i-(\omega\cdot x_i)-b) \\ &-\sum_{i=1}^{m}\alpha_i^*(\varepsilon+\xi_i^*-y_i+(\omega\cdot x_i)+b, \end{aligned}$$

where $\alpha^{(*)}=(\alpha_1,\alpha^*,\dots,\alpha_m,\alpha_m^*)^T$, and $\lambda^{(*)}=(\lambda_1,\lambda^*,\dots,\lambda_m,\lambda_m^*)^T$ are Lagrange multiplier vectors. We can then obtain the dual problem

$$\begin{aligned} \underset{\alpha^{(*)}\in R^{2m}}{\text{minimize}} \quad & \tfrac{1}{2}\textstyle\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j) \\ & +\varepsilon\textstyle\sum_{i=1}^{m}(\alpha_i^*+\alpha_i)-\sum_{i=1}^{m}y_i(\alpha_i^*-\alpha_i) \\ & \text{subject to} \quad \textstyle\sum_{i=1}^{m}(\alpha_i^*+\alpha_i)=0, \\ & 0\le\alpha_i^{(*)}\le C,\ i=1,\dots,m, \end{aligned} \quad (3)$$

On computing the dual problem, Equation (3), we now obtain the solutions of problems, Equation (2) as

$$\omega=\sum_{i=1}^{m}(\alpha_i^*-\alpha_i)x_i, \quad (4)$$

for this, select an element of α_i^*, $\alpha_j\in(0,C)$ or $\alpha_i^*\in(0,C)$ and subsequently solve for

$$b=y_i-\sum_{i=1}^{m}(\alpha_i^*-\alpha_i)(x_i\cdot x_j)+\varepsilon, \quad (5)$$

or

$$b=y_r-\sum_{i=1}^{m}(\alpha_i^*-\alpha_i)(x_i\cdot x_r)-\varepsilon. \quad (6)$$

Therefore, a point $x\in R$ has a corresponding y value given by

$$y=(\omega\cdot x)+b=\sum_{i=1}^{m}(\alpha_i^*-\alpha_i)(x_i\cdot x)+b. \quad (7)$$

Nonlinear problem

The linear optimization problem is now transformed into a nonlinear problem by introduction of a kernel function, $\varkappa$. In this case, it is observed that the inner products are present in the dual problem , Equation (3).Thus, the kernel functions can be directly applied and the other components are unchanged.

$$\begin{aligned} \underset{\alpha^{(*)}\in R^{2m}}{\text{minimize}} \quad & \tfrac{1}{2}\textstyle\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)+\varkappa(x_i,x_j) \\ & +\varepsilon\textstyle\sum_{i=1}^{m}(\alpha_i^*+\alpha_i)-\sum_{i=1}^{m}y_i(\alpha_i^*-\alpha_i), \\ & \text{subject to} \quad \textstyle\sum_{i=1}^{m}(\alpha_i^*+\alpha_i)=0, \\ & 0\le\alpha_i^{(*)}\le C,\ i=1,\dots,m. \end{aligned} \quad (8)$$

The conclusions made here are the same as those for the linear problem. Only that for the linear case the inner product (x,x') is considered in place of the kernel function $\varkappa(x,x')$.

5 Terms employed in the SVR framework

- **Kernel**: The function used to map a lower dimensional data into a higher dimensional data.
- **Hyper Plane**: In SVM, it is the separation line between the data classes. However, in the SVR framework, it is the line that will assist in predicting the continuous value or target value.
- **Boundary line**: In SVM there are two lines other than Hyper Plane which creates a margin . The support vectors can be on or outside the Boundary lines or outside it. This boundary line separates the two classes.
- **Support vectors**: These are the data points which are nearer to the boundary. The distance of the points is minimum or less.

5.1 Estimation Intuition

We briefly discuss the intuitions underlying the confidence of predictions from SVR. Consider a logistic regression, where the probability $p(y = 1/x; \theta)$ is modeled by $h_\theta(x) = g(\theta^T x)$. Then 1 can be predicted if and only if $h_\theta(x) \geq 0.5$ or equivalently, if and only if $\theta^T x \geq 0$. Again take a positive training example $(y = 1)$. The larger $\theta^T x$ the larger is $h_\theta(x) = p(y = 1/x; w, b)$, so the higher our degree of confidence that the label is 1. In this respect, we say informally that our prediction is a confident one that $y = 1$ if $\theta^T x >> 0$. Furthermore, the regression model would give a confident prediction for $y = 0$ where if $\theta^T x << 0$. In another intuition, consider Figure 1 in this, the x's represent positive training examples, and the o's are the negative training examples.A decision boundary (which is the line given by $\theta^T x = 0$, and is called the separating hyperplane) is also shown, and the three points have labeled A,B and C. For a comprehensive discussion of Figure 1, see [22]

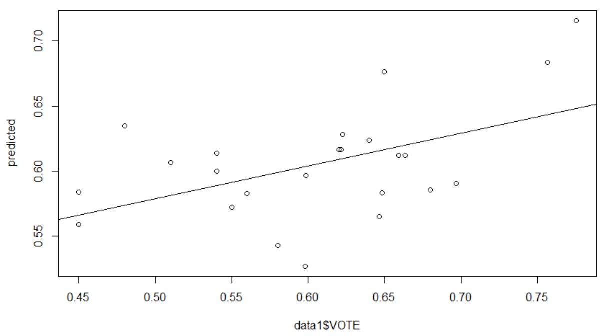

Fig. 1: SVR graph

In the SVR approach, the entire data set is split into two: The training set on which the regression is performed and the test set on which cross validation is done. In our study, the same procedure adheres to using the default set parameters.

6 Results

6.1 Beta model parameter optimization and interpretation

We consider prediction of the South African presidential elections, taking vote proportions (percentages) of the incumbent party (ANC) as the dependent variables and macroeconomic indicators, i.e. GDP, CPI, unemployment rate and GNP as the predictor variables. A data set for the relevant variables which spans the post apartheid era 1994 to 2017 is used.

Having used the `betareg` function in R, we had the following model:

$$logit(y_i) = \beta_o + \sum_{i=1}^{n} \beta_i x_i,$$

where $logit(y_i)$ is the log-odds. Thus the beta coefficients that `betareg` returns are the additional increase (or decrease for negative beta) in the log-odds of our response variable (Votes),
Given that we are interested in interpreting our betas on the probability scale (i.e. on the interval (0,1)), and having our beta coefficients all we require to do is to change the response, as

$$logit(y_i) = \beta_o + \sum_{i=1}^{n} \beta_i x_i \Rightarrow y_i = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i x_i}}.$$

Table 1: Beta model results

	coef.Estmates	Std. Error	z value	$Pr(>\|z\|)$
Intercept	2.561	1.118	2.290	0.022
GDP	-0.001	0.001	-2.685	0.007
CPI	0.067	0.070	0.963	0.335
UNEMPLOYEMENT RATE	-0.055	0.051	-1.063	0.287
GNP	-0.001	0.001	-0.481	0.630

The beta distribution was used to model incumbent party vote proportions, and the estimated model is given by

$$\hat{Vote} = 2.561 - 0.001GDP + 0.067CPI - 0.055UNEMPLYT - 0.001GNP \quad (9)$$

Model interpretation: We focus our interpretation of the model estimates on the odds rather than the log odds for simplicity. This is done by exponentiating the model coefficients in Table 1 and interpreting the result (odds) in percentage. The odds are given by probability of success divided by the probability of failure $= \frac{p}{1-p}$, where p is the probability of success.
From Equation (9), for a unit change in the South African real GDP, unemployment and GNP we expect to see a 0.1%, 5.5% and 0.1% decrease in the odds of the incumbent party votes respectively. Again, for a unit change in the CPI we expect to observe a 6.7% increase in the odds of the incumbent party votes.
OR
For a unit increase in the South African GDP, we expect to see 0.1% decrease in the odds of the incumbent party vote.
Note: if $\beta < 0$, changes occur in odds ratio (OR)= $(1 - e^{\beta}) \times 100$, and if $\beta > 0$, change in odds ratio is $(e^{\beta} - 1) \times$

To fully examine the performance of the beta distribution in predicting the incumbent presidential party vote, more statistical measures were carried out, See Table 2.

Table 2: presidential vote predictions with the Beta model

Year	Observed Vote %	Predicted Vote %	Prediction error
1994	62.25	62.85	0.60
1995	65.00	67.61	2.61
1996	64.00	62.37	-1.63
1997	62.00	61.69	-0.31
1998	54.00	61.39	7.39
1999	66.35	61.21	-5.14
2000	68.00	58.56	-9.44
2001	51.00	60.66	9.66
2002	56.00	58.26	2.26
2003	54.00	60.00	6.00
2004	69.69	59.05	-10.64
2005	64.67	56.52	-8.15
2006	58.00	54.28	-3.72
2007	59.80	52.70	-7.09
2008	45.00	55.93	10.93
2009	65.90	61.24	-4.66
2010	44.98	58.40	13.42
2011	55.00	57.23	2.23
2012	64.85	58.34	-6.50
2013	59.87	59.65	-0.22
2014	62.15	61.64	-0.51
2015	48.00	63.50	15.50
2016	75.65	68.37	-7.28
2017	77.57	71.57	-6.00

Table 2 indicates the observed and predicated vote percentages of South African presidential elections. It should be noted that when elections did not take place, because they are not held annually, the percentages of intention to vote the incumbent party were used. By visual inspection, see Table 2, the predicted votes do not appear good enough in some instances, though relatively good for selected years. The prediction errors are also shown, but are not good enough on the overall because some years have big errors. However, from the model output the mean square error (MSE) = 1.216 and Pseudo $R^2 = 28.51\%$. At this point we excuse ourselves making much sense of the pseudo R^2 values because no perfect model exists and the existence of measurement error means that the maximum that a perfect model can "explain" is not 100%.
We consider a data visualization of the scatter plot of the predicted versus actual votes, Figure **??**.

From the scatter plot, Figure **??**, we can say something about other issues related to the fitted model, Equation (9). It is clear that the residuals are heteroskedastic, implying that the variance of the error is inconstant across various levels of the vote proportions. Consequently the standard errors of the regression coefficients in Table 1 are unreliable and may be understated. Therefore, the statistical significance of the considered macroeconomic indicators in Equation (9) could have been overstated. They may not be statistically significant due to the heteroskedastic issue, though we may be unable to exactly ascertain this.

6.2 SVR model parameter optimization and interpretation

Considering the training of the SVR models, one kernel parameter γ, and two kernel-independent parameters ε and C are required to be defined, Equation (2). The basis for determining these parameters is to build a model that best fits our data. However, in this research, values recommended by [21] are considered because there are no known standard approaches for dictating optimal SVR model parameters [23]. Hence, it is more of `an art than a science` to determine appropriate values for these parameters. Reviewed literature on SVR indicates that these parameters determined one at a time after letting other parameters constant, and subsequently allowing the parameter take on different values, then identifying the value that seems to align with the best model performance assessed by cross-validation [24].

In Table 3, we indicate the parameters (ε and C) recommended by [21] and our optimized parameters (ε, C and γ) of the SVR model constructed for the incumbent presidential party vote.

Table 3: SVR-Incumbent Vote (SVR-IV) model parameters

Model	$\varepsilon-$ recommended	$C-$ recommended	$C-$ optimized
SVR-IV-1	0.0475	10.34	1
SVR-IV-2	0.0152	10.00	1

The SVR-IV-1 and SVR-IV-2 models constructed for the incumbent party vote demonstrated consistently good performance for the `esp-Radial` models with `MSE` = 0.0072 and $R^2 = 23.54\%$, `MSE` = 0.0060 and $R^2 = 21.48\%$ respectively using the recommended parameters as shown in Table 3.
We now compare the prediction capabilities of the two models (beta and SVR) on the test data. Recall that model validation for the SVR was performed on the test data and the years randomly selected for the test data are specifically extracted out for the beta model from Table 2. Table 4 indicates the comparison of vote predictions using both the SVR and beta regression models. The best model for the South African presidential vote election predictions has been built via SVR using the recommended parameters as earlier mentioned.

Using the best model (SVR-radial), ascertained using the `MSE` result, the forecast for the incumbent vote based on quarterly macroeconomic indicators;
GDP = 3186.5, CPI = 112.9, UNEMPLYT = 24.7, GNP = 3096.7 is 60.32%. In other words, it is envisioned that the ANC party will win the presidential forthcoming elections with a percentage of 60.32%.

Table 4: Comparison of vote predictions using SVR and the Beta models

Year	$VOTE$	$S_{Predicted}$	S_{error}	$B_{Predicted}$	B_{error}
1994	0.6225	0.6437	0.0212	0.6285	0.0060
2004	0.6969	0.5779	-0.1190	0.5905	-0.1064
2007	0.5980	0.5736	-0.0244	0.527	-0.0710
2010	0.4498	0.6136	0.1638	0.584	0.1342
2011	0.5500	0.6109	0.0609	0.5723	0.0223
2012	0.6485	0.6061	-0.0423	0.5834	-0.0651
2014	0.6215	0.5973	-0.0242	0.6164	-0.0051
2015	0.4800	0.5978	0.1178	0.635	0.1550
2016	0.7565	0.6002	-0.1564	0.6837	-0.0728
2017	0.7757	0.6022	-0.1735	0.7157	-0.0600

Note: S=SVR B=Beta

7 Conclusion

The macroeconomic indicators, i.e. gross national product, consumer price index, unemployment rate and gross national product, that could explain presidential voting have been explored with a view of achieving a prediction model for presidential elections. Two models i.e. the support vector regression and the beta distribution models. Based on the MSE = 0.006 and 1.216 for both models respectively, the SVR turns out to be the best model and predicts the incumbent party (ANC) vote with a 60.32% majority win for the South African presidential elections held on 8th March 2019. The predicted vote percentage is compares with the actual (obtained) percentage of 57.5%. Error in the predicted vote percentage could be due to the fact that some important macroeconomic indicators were not considered in this study. However, predictions from the SVR have been rather more accurate. Despite this admirable evidence, one should not be unduly optimistic about SVR performance because it relies on the past to predict the future, patently a precarious business. In this research, the past is the post apartheid period (1994-2017) which could be considered ‘stable’ with relatively little political and economic disturbances. However, the two approaches provide a scientific basis for presidential vote prediction.

Acknowledgment

This research has been partly supported by the Human Sciences Research Council (HSRC) of South Africa, Faculty of Health and Applied Sciences (Namibia University of Science and Technology), Department of Economics and Statistics, and Faculty of Economics and Management Sciences, Kabale University.
We extend our sincere thanks to the two anonymous reviewers who provided insightful comments that presented a distinctive paper.

References

[1] S. Ferrari and F. Cribari-Neto, Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics **31**:7, 799-815 (2004). DOI: 10.1080/0266476042000214501
[2] D.T Bui, T.A Tuan, H. Klempe, B. Pradhan, I. Revhaug, Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree, Landslides **13**, (2)361-378 (2016).
[3] H.C Wu, M.J Ho, T.D Lee, Travel-Time Prediction With Support Vector Regression, IEEE Transactions on Intelligent, Transportation Systems **5**, (4)276-281 (2004).
[4] R.S Pindyck and L.D Rubinfeld, Econometric Models and Economic Forecasts, New York: McGraw-Hill, 1976.
[5] J.S Long, Regression Models for Categorical and Limited Dependent Variables, Sage Publishing, (1997).
[6] G.M Karen, https://www.theanalysisfactor.com/proportions-as-dependent-variable-in-regression-which-type-of-model/,[Accessed on: 5 April 2019].
[7] A.Linden, W.M Trochim, J.L Adams, Evaluating program effectiveness using the regression point displacement design, Evaluation & the Health Professions **29**, (4)407-423 (2006).
[8] A. Agresti, An introduction to categorical data analysis, Wiley, (2018).
[9] S.S Mangiafico, Summary and Analysis of Extension Program Evaluation in R, version 1.15.0. rcompanion.org/handbook/. (Pdf version: rcompanion.org/documents/RHandbookProgramEvaluation.pdf.), (2016).
[10] F. Cribari-Neto and A. Zeileis, Beta regression in R, Journal of Statistical: Software **34**, 1–24 (2010).
[11] M. Smithson, and J. Verkuilen, A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, Psychol Methods **11**, 54–71 (2006).
[12] C.H Lima, H.H Kwon, J.Y Kim, A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate, Journal of Hydrology **540**, pp.744-756 (2016).
[13] C.E.B Owen, Parameter estimation for the beta distribution, Masters Thesis: Brigham Young University - Provo, (2008).
[14] S. Kotz, Handbook of Beta distribution and its applications, Journal of International Biometric Society **62**, (1)309-310 (2006)
[15] K. Krishnamoorthy, Handbook of statistical distributions with applications, Chapman and Hall/CRC, (2016).
[16] L.J Bain and M. Engelhardt, Introduction to Probability and Mathematical Statistics, Duxbury Press, (1991).
[17] V. Cherkassky, and F. Mulier, Learning from data: Concepts, theory,and methods, New York: Wiley, (1998).
[18] V. Cherkassky, and Y. Ma, Comparison of model selection for regression, Neural Computation **15**, (7): 1691-1714 (2003).
[19] V. Cherkassky, and F. Mulier, Statistical learning theory, New York: Wiley, (1998).

[20] V. Cherkassky, and F. Mulier, The nature of statistical learning theory, (2nd ed), Berlin, Springer, (1998).
[21] V. Cherkassky, and Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Network **17**, 113-126 (2004).
[22] D. Basak, S. Pal, C.D Patranabis, Support Vector Regression, Neural Information Processing **11**, 203-226 (2007).
[23] L. Tongbin, L. Wen, X. Qiqi, Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models, BMC Bioinformatics **7**, 182-194 (2005).
[24] X.J Yao, A. Panaye, Z.D Hu, Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression, J Chem Inf Comp Sci 1251-1266 (2004).

C. R. Kikaw holds a PhD degree in Statistics. He is currently a lecturer in appllied statistics at Kabale University, Uganda. His research interests are in statistical modelling and estimation theory.

M. Nduku holds a PhD in Mathematical Technology from Tshwane University of Technology. Mercy has published several peer-reviewed publications and presented in conferences, her research interest is estimation methods with incomplete data or information and predictive modeling.

D. Ntirampeba has a PhD degree in Applied Statistics. He is a lecturer of Statistics at Namibia University of Science and Technology. His research interest is in non-linear mixed modelling, spatial and spatio-temporal modelling.

A. Ssematimba has a PhD degree in Applied Mathematics. He is a senior lecturer in Mathematics. His research interest is in mathematical modelling and population dynamics.