
Support Vector Regression and Artificial Neural Network Approaches: Case of Economic Growth in East Africa Community

Abraham Kipkosgei Lagat^{*}, Anthony Gichuhi Waititu, Anthony Kibera Wanjoya

Department of Basic and Applied Science, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

abraham.lagat.k@gmail.com (A. K. Lagat), agwaititu@gmail.com (A. G. Waititu), awanjoya@gmail.com (A. K. Wanjoya)

^{*}Corresponding author

To cite this article:

Abraham Kipkosgei Lagat, Anthony Gichuhi Waititu, Anthony Kibera Wanjoya. Support Vector Regression and Artificial Neural Network Approaches: Case of Economic Growth in East Africa Community. *American Journal of Theoretical and Applied Statistics*.

Vol. 7, No. 2, 2018, pp. 67-79. doi: 10.11648/j.ajtas.20180702.13

Received: September 1, 2017; **Accepted:** September 18, 2017; **Published:** March 16, 2018

Abstract: There has been increased interest of late on the application of nonlinear methods to economic and financial data due to their robustness in handling large and complex data. With increasingly complex ‘big data’, focus has shifted into use of robust techniques in analysis of data. Various nonlinear approaches have so far been established including support vector machine which is widely adapted in classification and regression problems. This research project applied support vector regression technique and neural network models in modeling and forecasting economic growth for the five countries in the East Africa Community including Kenya, Uganda, United Republic of Tanzania, Rwanda and Burundi. Data for the period 1990 to 2014 from World Bank databases was used for the research. Support vector model and neural network models were trained using the data for the 1990-2002 whereas the remaining data was used for prediction performance to determine the robustness of the two models on external datasets. The study revealed that specific country models had better performance compared to the combined model and that although the two models compared similarly under specific-country models, the neural network performed better in most countries. The study recommends the use of the two machine learning techniques in economic growth modeling. It also recommends that the performance be compared with the traditional econometric models but using countries with more data periods.

Keywords: Support Vector Regression, Artificial Neural Network, Machine Learning, Economic Growth, East Africa, Economic Data

1. Introduction

Economic growth is a complex process affected by a number of factors, and theory gives us no clear or single answer to the question about the right model specification. The number of variables collected in economic data have also grown enormously and become highly complex to be modeled with specified relationships only.

Of late there has been an increased interest on the application new statistical methods of nonparametric regression methods to big data. These methods are considered the most flexible approach in nonlinear modeling since they do not have defined assumptions that have restrictive application that require specific considerations.

Researchers have been utilizing economic theory to define the structural relations between economic growth and other variables. Besides, statistical approaches have been developed to identify correlation in historical tendencies such as the famous Solow-Swan model [1, 2] and Cobb- Douglas models [3]. Others include the gray theory method [4], autoregressive moving average (ARIMA) model, regression model and random walk model. However, due to various assumptions such as linearity, these models can hardly be reliable to be used in forecasting [5].

Recently, artificial neural networks (ANN) have had increasing attention in financial forecasting due to their nonlinear mapping capabilities and data processing characteristics. The application of markov switching models

on the US gross national product (GNP) data [6] enhanced consideration of applying nonlinear models which has led to the large techniques that have been developed [7] including the logistic regression, Support Vector Machine (SVM), artificial neural networks (ANN), Holts exponential smoothing and machine learning techniques. These methods have been applied to economic time series [8], weekly foreign exchange rates [9], daily foreign exchange rates [10] as well as bankruptcy prediction [11]. The techniques have also been applied in tourism and hospitality [12, 13], energy [14, 15] and meteorology [16].

SVM is one of the types of supervised learning techniques introduced in the early 1960s. It is in the family of generalized linear classifiers and which uses classification and regression machine learning theory to maximize predictive accuracy and avoiding over-fitting data simultaneously. Under SVM, rather than implementing the empirical risk minimization (ERM) principle to minimize the training error, it employs the structural risk minimization (SRM) principle to minimize an upper bound on the generalization error, and allow learning any training set without error.

Such models, having been introduced by Vapnik in the 1990's [17, 18], have proved to be effective and promising techniques for handling linear and nonlinear classification and regression [19, 20] due to their principle of maximal margin, dual theory and kernel trick. In contrast to other nonparametric methods such as Neural Networks, SVMs has become powerful tools to solve the problems and overcome some traditional difficulties such as over-fitting. It has performed exceptionally well in various fields of study. Specifically, SVR has been investigated and found to be the efficient, accurate and robust in handling complex analysis. This has seen the application of SVR surpass other machine learning approaches including the ANN.

This research applied SVR on the economic growth in the East Africa Member countries. There are three main measures for economic growth. These are the gross domestic product (GDP), gross domestic product on purchasing power parity (GDP PPP) and the human development index.

The most common parameter used to gauge economic growth is the gross domestic product (GDP) and its derived indicators such as the gross national product (GNP) and gross national income (GNI). Although this value is derived by dividing the gross domestic product by the population, it does not take into account the distribution amongst the population.

GDP PPP is gross domestic product converted to international dollars using purchasing power parity rates. The concept which is based on the law of one price was introduced by the School of Salamanca in the 16th century and developed by Gustav Cassel [21]. PPP exchange rates help to minimize misleading international comparisons that can arise with the use of market exchange rates.

The human development index (HDI) has been utilized by the United Nation Development Programme in its Human development reports in rating countries in respect to life expectancy, education and per capita GDP PPP.

The GDP PPP has been used as the measure of economic

growth in this research work. The research sought the possibility of developing an SVR model for economic analysis. In this research work, epsilon-SVR and a backward-propagated Artificial Neural Network (ANN) models were developed and the results compared both for the train and test data for country-specific as well as combined data.

2. Methodology

2.1. Research Data and Data Preparations

This research study used the official annual economic data from the World Bank and IMF websites. Data covering the period 2000 to 2014 was obtained from the World Bank databank (<http://databank.worldbank.org/data/databases.aspx>) and the 2016 data estimate from the IMF data portal (<https://www.imf.org/external/data.htm>). The study focused on the economic data of the 5 East Africa countries that make up East Africa Community. These are Kenya, Uganda, United Republic of Tanzania, Rwanda and Burundi.

The data from the World Bank was filtered on country criteria to ensure only data for the East Africa region covering the period 1990 to 2014. Variables contributing to economic growth selected for modeling. The physical capital, inflation, exchange rates, research and development investment, population growth, financial development, international trade are some of the variables included in economic growth models by the OECD [22]. Solow [1] and Swan [2] models have a real income output with labour, capital and knowledge variables as the independent variables.

Factors influencing economic growth can be categorized into human capital investment, natural resources, physical capital investment and entrepreneurship whereas the main dependent variable measuring economic growth is the gross domestic product per capita PPP (GDP_PPP).

Scaling transformation of the all the independent variables into [0,1] was done to avoid domination of features with high numeric range values. The data was divided into training and test datasets.

2.2. Support Vector Regression Model

Given training dataset, $D = \{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \mathbb{R}, i = 1, 2, \dots, n\}$. That is, X_i is a multivariate input consisting of all the independent variables, y_i is the corresponding scalar output and n is the number of the training samples. The support vector regression is defined as,

$$y_i = f(x, \omega) = \sum_i^n \omega \cdot \Phi_i(x) + b \quad (1)$$

where ω is the weight vector corresponding to $\Phi_i(x)$, $\Phi_i(x)$ nonlinear transformations mapping function, and b is a constant threshold. The parameters ω and b need to be estimated.

The variables

y_i - a vector of values for the gross domestic product per capita PPP (GDP_PPP) for each of the East Africa Community members for the period 2000-2014

x - matrix of variables consisting of the human capital

investment, natural resources, physical capital investment and entrepreneurship corresponding to y_i

2.3. Parameter Estimation

Flatness in the regression model means that one is seeking a small ω . The values of ω and b can be estimated by minimizing the following formula based on structural risk minimization principle:

$$R(C) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n L_\varepsilon(f(x_i, \omega)) \quad (2)$$

Where ε is predetermined value, $L_\varepsilon(x, \omega)$ is the empirical error measured by ε -insensitive loss function. Under SVR The empirical error is defined as,

$$L_\varepsilon(f(x_i, \omega)) = \begin{cases} 0, & \text{if } |y_i - f(x_i, \omega)| < \varepsilon \\ |y_i - f(x_i, \omega)|, & \text{otherwise} \end{cases} \quad (3)$$

Thus to minimize the norm $\|\omega\|^2 = \langle \omega, \omega \rangle$, involves solving a convex optimization problem:

$$\text{Minimize, } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4)$$

subject,

$$\begin{cases} y_i - f(x_i, \omega) - b \leq \varepsilon + \xi_i \\ f(x_i, \omega) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (5)$$

Where ξ_i and ξ_i^* are slack variables introduced due to the error in fitting in the optimization problem. They represent the distances from actual value to the corresponding boundary value of ε -tube as shown in figure 1 below.

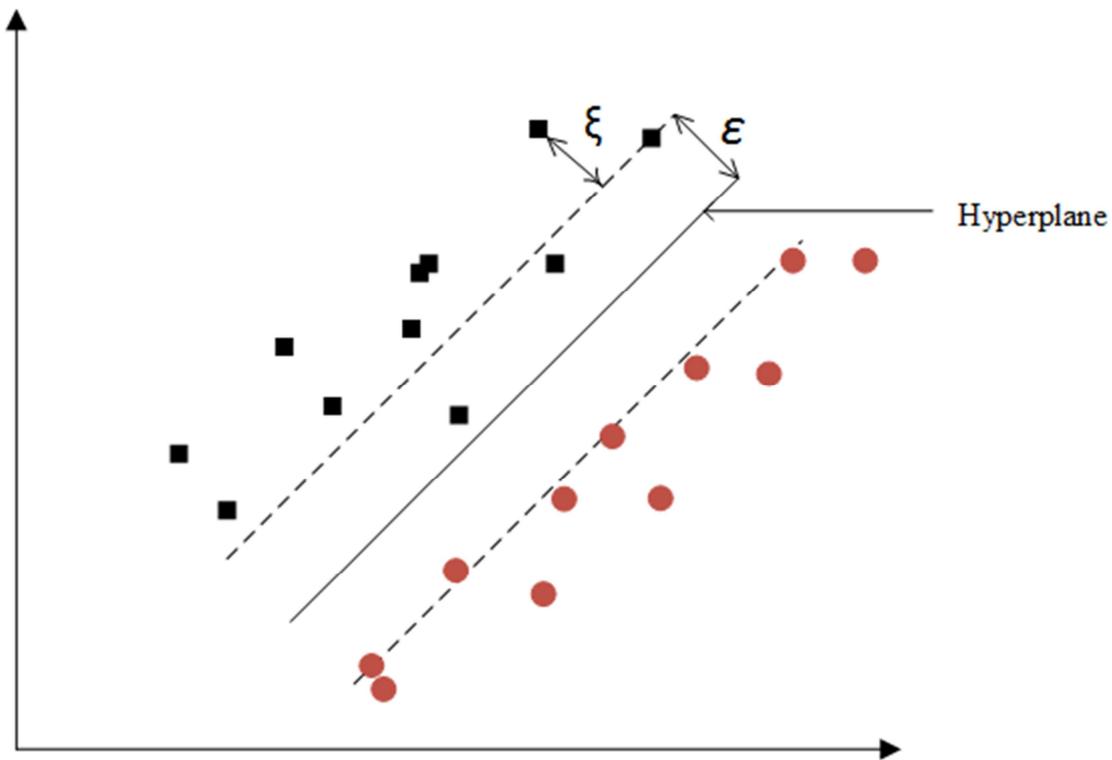


Figure 1. Setup of a linear SVR.

The primal problem can also be transformed into a dual problem and its solution is given by

$$\begin{aligned} f(x) &= \sum_{i=1}^{n_{sv}} (\alpha_i - \alpha_i^*) K(x_i, x) \\ \text{s.t. } 0 &\leq \alpha_i^* \leq C, 0 \leq \alpha_i \leq C, \end{aligned} \quad (6)$$

where n_{sv} is the number of Support Vectors (SVs) and $K(x_i, x)$ is the kernel function.

The constant $C > 0$ determines the trade-off between the fitness of SVR function and the amount up to which deviations larger than ε are tolerated. Optimal choice of regularization parameter C can be derived from standard parameterization of SVM solution.

Finally, estimation of b done by exploiting the so called Karush Kuhn Tucker (KKT) conditions [23, 24] which state that at the optimal solution the product between dual variables and constraints has to vanish. In SVR this means,

$$\alpha_i (\varepsilon + \xi_i - y_i + \langle \omega, x_i \rangle + b) = 0, \quad (7)$$

$$\alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle \omega, x_i \rangle + b) = 0, \quad (8)$$

and

$$(C - \alpha_i) \xi_i = 0, \quad (9)$$

$$(C - \alpha_i^*) \xi_i^* = 0, \quad (10)$$

From this, several useful conclusions can be deduced.

Firstly only samples (x_i, y_i) with corresponding $\alpha_i^* = C$ lie outside the ε -insensitive tube. Secondly $\alpha_i \alpha_i^* = 0$. That is there can never be a set of dual variables $\alpha_i \alpha_i^*$ which are both simultaneously nonzero as this would require nonzero slacks in both directions. For $\alpha_i^* \in (0,1)$, $\xi_i^* = 0$ while the second term vanishes. Hence b can be computed as follows,

$$b = y_i - \langle \omega, x_i \rangle - \varepsilon, \text{ for } \alpha_i \in (0,1) \quad (11)$$

$$b = y_i - \langle \omega, x_i \rangle + \varepsilon, \text{ for } \alpha_i^* \in (0,1) \quad (12)$$

2.4. Comparative Performance with Artificial Neural Network

The study compared the performance of the support vector regression with the artificial neural network using root mean square error (RMSE).

2.5. Artificial Neural Network Procedure

Consider a supervised learning problem with labeled training examples $(x^{(i)}, y^{(i)})$. Neural networks give a way of defining a complex, non-linear form of hypotheses $h_{w,b}(x)$, with parameters W, b that to fit the data.

The diagram below describes the simplest possible neural network comprising of a single neuron.

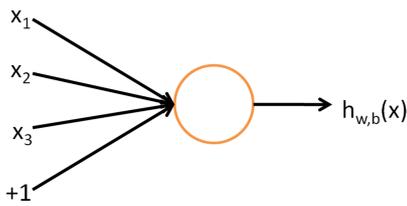


Figure 2. Neuron illustration.

This neuron is a computational unit that takes as input x_1, x_2, x_3 (and a +1 intercept term), and outputs

$$h_W = f(\sum_{i=1}^3 W_i x_i + b) \quad (13)$$

where $f: \mathbb{R} \rightarrow \mathbb{R}$ is called the activation function.

A neural network is put together by combining many of our simple neurons, so that the output of a neuron can be the input of another.

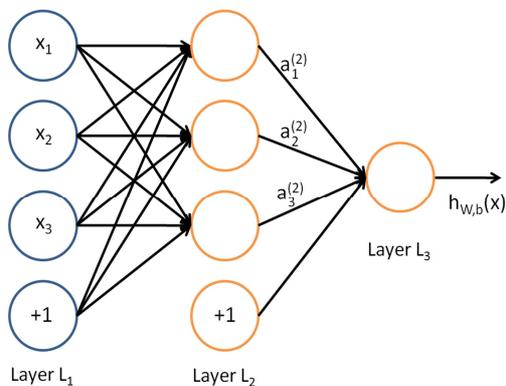


Figure 3. Neural network model.

In figure 4 above, the circles labeled “+1” are called bias

units, and correspond to the intercept term. The leftmost layer of the network is called the input layer (L_1), and the rightmost layer the output layer (L_3). The middle layer (L_2) of nodes is called the hidden layer, because its values are not observed in the training set. This neural network has 3 input units 3 hidden units, and 1 output unit.

2.6. Back Propagation

Let $\delta^{(l+1)}$ be the error term for the $(l+1)$ -st layer in the network with a cost function $J(W, b; x, y)$ where (W, b) are the parameters and (x, y) are the training data and label pairs. If the l -th layer is densely connected to the $(l+1)$ -st layer, then the error for the l -th layer is computed as

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) * f'(z^{(l)}) \quad (14)$$

and the gradients are

$$\nabla_W^{(l)} J(W, b; x, y) = \delta^{(l+1)} (a^{(l)})^T \quad (15)$$

$$\nabla_b^{(l)} J(W, b; x, y) = \delta^{(l+1)}. \quad (16)$$

If the l -th layer is a convolutional and subsampling layer then the error is propagated through as

$$\delta_k^{(l)} = \text{upsample} \left((W_k^{(l)})^T \delta^{(l+1)} \right) * f'(z_k^{(l)}) \quad (17)$$

Where k indexes the filter number and $f'(z_k^{(l)})$ is the derivative of the activation function. The upsample operation has to propagate the error through the pooling layer by calculating the error with respect to each unit incoming to the pooling layer. In max pooling the unit which was chosen as the max receives all the error since very small changes in input would perturb the result only through that unit.

Finally, calculation of the gradient with respect to the filter maps relies on the border handling convolution operation again and flip the error matrix $\delta^{(l)k}$ the same way, the layers is flipped in the convolutional layer.

$$\nabla_W^{(l)} J(W, b; x, y) = \sum_{i=1}^m (a^{(l)}_i) * \text{rot90}(\delta^{(l+1)}_k, 2) \quad (18)$$

$$\nabla_b^{(l)k} J(W, b; x, y) = \sum_{i=1}^m (a^{(l)}_i)_{a,b} \quad (19)$$

Where $a^{(l)}$ is the input to the l -th layer, and $a^{(1)}$ is the input image. The operation $(a^{(l)}_i) * \delta^{(l+1)}_k$ is the valid convolution between i -th input in the l -th layer and the error with respect to the k -th filter.

2.7. Model performance and forecasting

The performance of the model will be assessed on the validation dataset using both the mean square error (MSE).

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (20)$$

Where y_i is the i^{th} observed value, \hat{y}_i is the i^{th} fitted value and n is the total number of validation data.

The developed SVR model was used to predict outputs of the given inputs in the test data.

The predicted economic growth indicator per capita PPP (lnGDP_PPP) will, therefore, be:

$$\hat{y}_i = \sum_{i=1}^n \hat{\omega} \cdot \Phi_i(x) + \hat{b}, \tag{21}$$

Where \hat{y}_i is the predicted lnGDP_PPP values, x matrix data for 2016, $\hat{\omega}$ and \hat{b} are the parameters estimated and optimized in the previous sections.

The analysis applied various packages including e1071, caret and neuralnet packages in R statistical R Version 3.1.

3. Empirical Results

3.1. Pre-processing and variable selection

To enable utilization of both models to be compared (Neural network and support vector regression), the variable importance was determined using model independent metrics whereby each variable is evaluated by filtering.

In this case, the relationship between each possible independent variable with the dependent variable was calculated. Variable selection for the modeling was done through principal component analysis (PCA) with scaling.

A cutoff eigenvalue of 1 was used to select the number of principal components reducing them to only four ($PC1=3.02$, $PC2=2.83$, $PC3=2.03$ and $PC4=1.27$) which explains over 76% of the variance. The dependent variable

is correlated to the first two principal components ($PCA1_{ry}=0.474$, $PCA2_{ry}=0.690$) and 8 of the independent variables a correlation of at least 0.4 with the either of the first two. Further analysis was carried out to remove variables with almost constant values which could affect the models. Variables with over 90% frequency ration were included for further analysis.

Table 1. PCA results.

	Eigen value	percentage of variance	cumulative percentage of variance
comp 1	3.023	25.193	25.193
comp 2	2.835	23.621	48.814
comp 3	2.034	16.948	65.762
comp 4	1.270	10.585	76.346
comp 5	0.917	7.645	83.991
comp 6	0.610	5.084	89.075
comp 7	0.472	3.935	93.011
comp 8	0.378	3.154	96.164
comp 9	0.157	1.310	97.474

3.2. Summary Statistics

The distribution of all the variables used are in figure 5 below. the GDP growth has been fluctuating around zero for all the East African countries. Around 1994, Rwanda had the lowest GDP growth ever experienced by any other country. This was due to the effect of genocide which has also been reported elsewhere [25, 26].

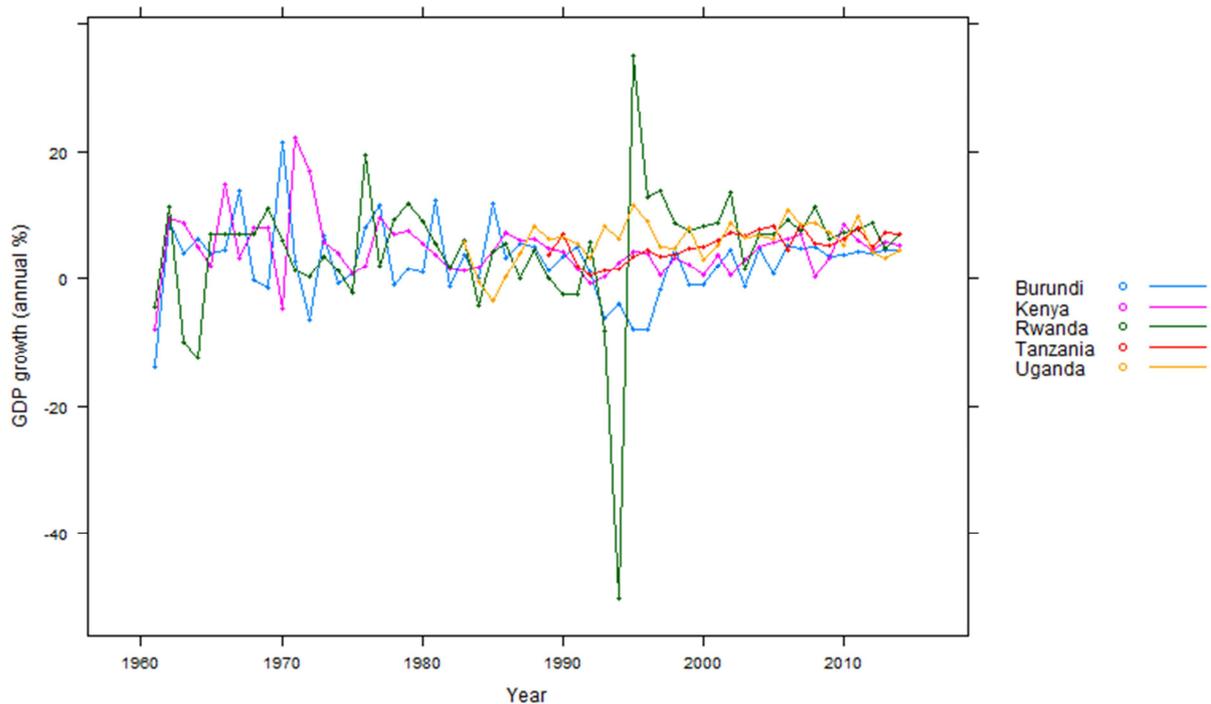


Figure 4. GDP for the different member countries.

On the other hand, Burundi has had the lowest economic growth compared to the other members of the community from the mid-90s to date except towards 2010 when Kenya had the lowest economic growth which has been attributed to the 2007/2008 post-elections violence.

3.3. SVR Modeling and hyper-parameter optimization

The optimization of the SVR parameters was carried out using the standard grid search technique having been utilized by other authors [27]. In this way, the best combination of the

epsilon, cost and gamma parameters was chosen based on the error obtained from the training data. A 5-fold cross validation

method was included in this process. The optimized SVR hyperparameters per country is presented in table 2 below.

Table 2. Optimized SVR parameters.

Parameter	Gamma (γ)	Cost (C)	Epsilon (ϵ)	Support Vectors	Hidden Nodes
Burundi	0.05	5	0.001	13	4
Kenya	0.85	5	0.85	7	3
Rwanda	0.05	10	0.25	18	1
Tanzania	0.05	5	0.45	10	1
Uganda	0.001	30	0.001	5	1
Combined	0.05	55	0.05	17	3

All the optimal hyperparameters varied across the countries. Generally, Kenya ($\gamma=0.85$) and the combined model ($\gamma=0.05$) had higher gamma parameters compared to Burundi ($\gamma=0.05$), Uganda ($\gamma=0.001$), Rwanda ($\gamma=0.05$) and Tanzania ($\gamma=0.005$). These two SVR models would be more nonlinear as compared to the rest. The cost parameters were lower for Burundi, Kenya and Tanzania ($C=5$) which indicates lower penalties to support vectors as compared to those for Uganda ($C=30$), Rwanda ($C=10$) and the combined mode ($C=55$).

The ϵ -tube separating hyperplane for Kenyan model ($\epsilon=0.85$) is approximately as twice for Tanzania ($\epsilon=0.45$) and about 3.5 times for Rwanda ($\epsilon=0.25$). SVR model for Burundi ($\epsilon=0.001$), Uganda ($\epsilon=0.001$) and the combined countries ($\epsilon=0.05$).

For the neural network model, all the specific models performed better than the combined model (MSE=0.211) which yield about 10 times poorer than the specific model for Kenya (MSE=0.02), 5 times that of Tanzania (MSE=0.042), Rwanda (MSE train=0.064) and Uganda (MSE=0.048). The specific model for Burundi had the lowest training error (MSE=0.006).

3.4. Comparative Performance of SVR and Neural Network

Based on the training models, table 3 below are the results of the SVR and neural network models. The results presented are the fitted values obtained using the respective models as well as the observed values for the period 1990 – 2002.

Table 3. Performance of SVR and neural network on train data.

	Country-Specific model		Combined model	
	SVR	NNET	SVR	NNET
Burundi	8.154	2.628	5.008	35.085
Kenya	2.463	2.896	2.217	4.051
Rwanda*	2.471	4.241	1.089	3.142
Tanzania	1.415	0.344	1.293	4.728
Uganda	5.395	5.411	5.852	17.156

*Excluded 1994-1995 data

It is clear that the country-specific models have better fits compared to the combined model.

The mean square error from the test data is an indication of the generalization of the models to external data. Using the combined model, the mean square error for the SVR model is smaller across all the countries. As for the country-specific model, neural network outperformed SVR in Burundi (MSE SVR = 2.628, MSE NNET=8.154) and Tanzania (MSE SVR = 1.415, MSE NNET=0.344). The neural network is better to

the SVR in the combined model.

3.5. Prediction results

Based on the test data covering the period 2003 - 2014, below are the results both for the SVR and neural network. In undertaking the prediction, the poor performance of the combined models both in SVR and neural network was taken into account and therefore only performance of the country-specific SVR and neural network models was tested.

Table 4. Performance of country-specific SVR and neural network on test data.

	Country-Specific model		Combined model	
	SVR	NNET	SVR	NNET
Burundi	22.715	7.577	29.519	3.719
Kenya	10.143	12.980	20.973	7.753
Rwanda*	6.905	6.927	23.987	20.636
Tanzania	2.918	9.361	32.891	14.019
Uganda	4.864	4.791	37.102	17.977

Based on the table 4 above, except the SVR model for the Burundi which performed extremely poor compared to the neural network, the two models were generally robust outside the training data.

4. Conclusions and Recommendations

This work sought to establish the application of support vector regression in modeling economic growth using the East Africa Community data obtained from the World Bank, compare its performance to the neural network model and determine their robustness in forecasting.

The comparative performance of the two models based on their mean square errors indicated similar performance. However, the neural network model was better compared to the support vector regression for Tanzania and Burundi under the specific-country models while it performed better in combined model for all countries.

The robustness of the models in using external test data for the period 2003-2014 showed that the two models compared similar except for Burundi in which neural network highly outperformed the SVR and vice-versa for Tanzania. Using the combined model, neural network outperformed SVR across all countries.

Appendix

Appendix I: Optimized Support Vector Regression Parameters

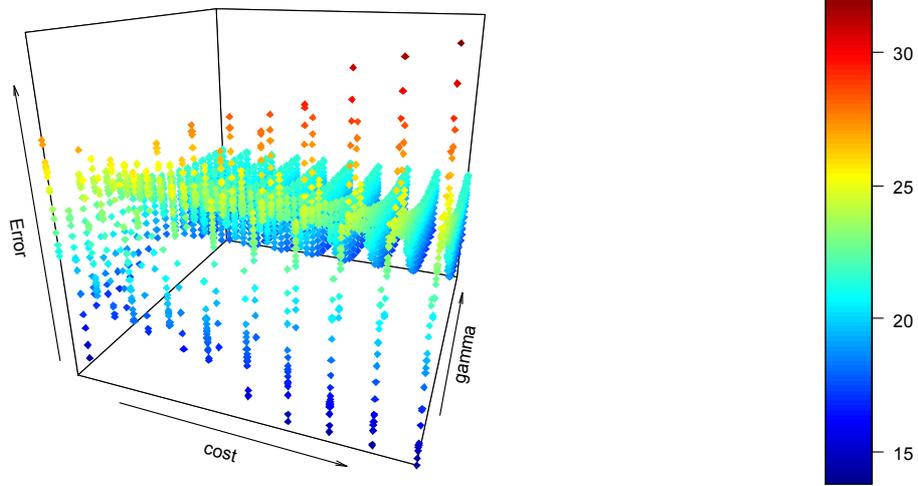


Figure 5. Burundi SVR optimization.

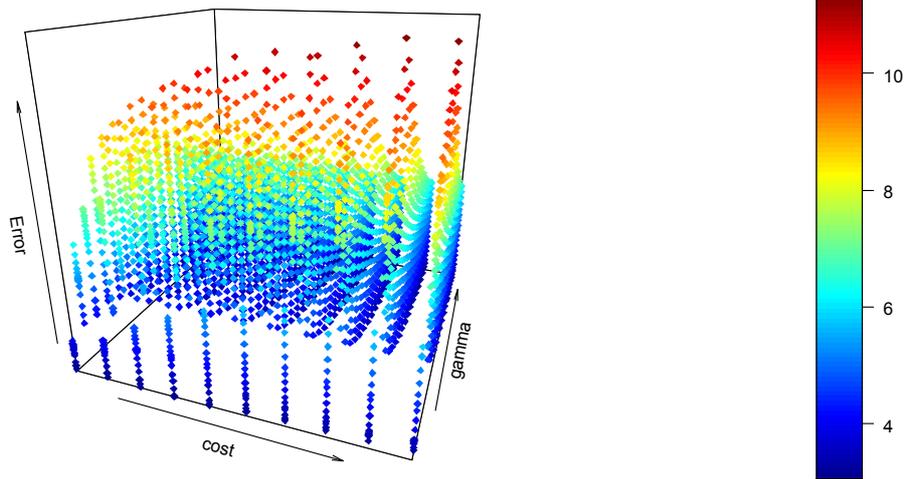


Figure 6. Kenya SVR optimization.

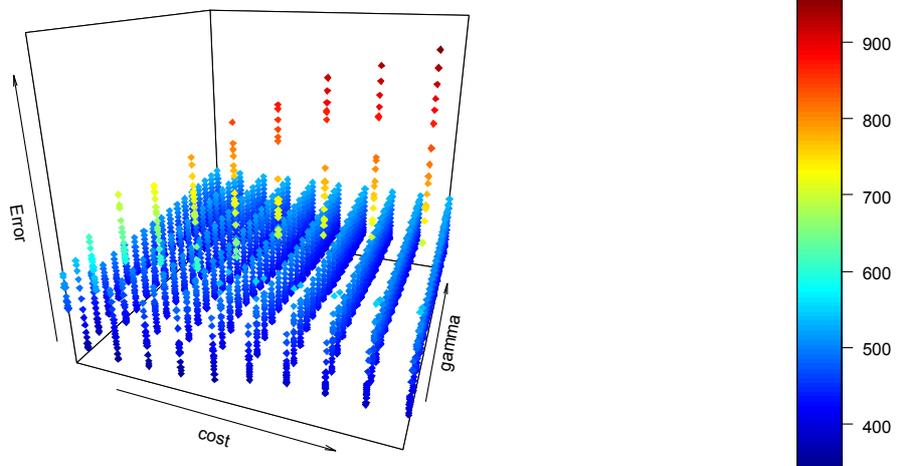


Figure 7. Rwanda SVR optimization.

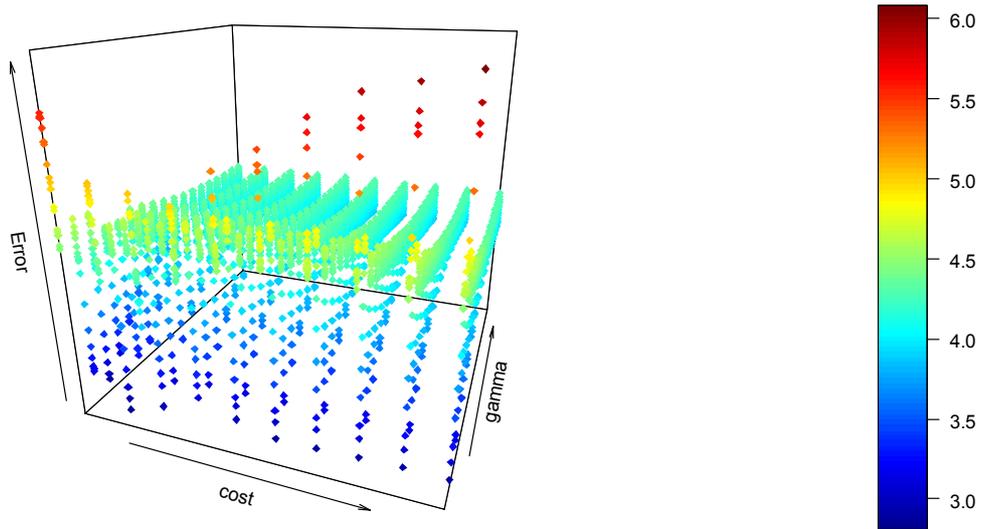


Figure 8. Tanzania SVR optimization.

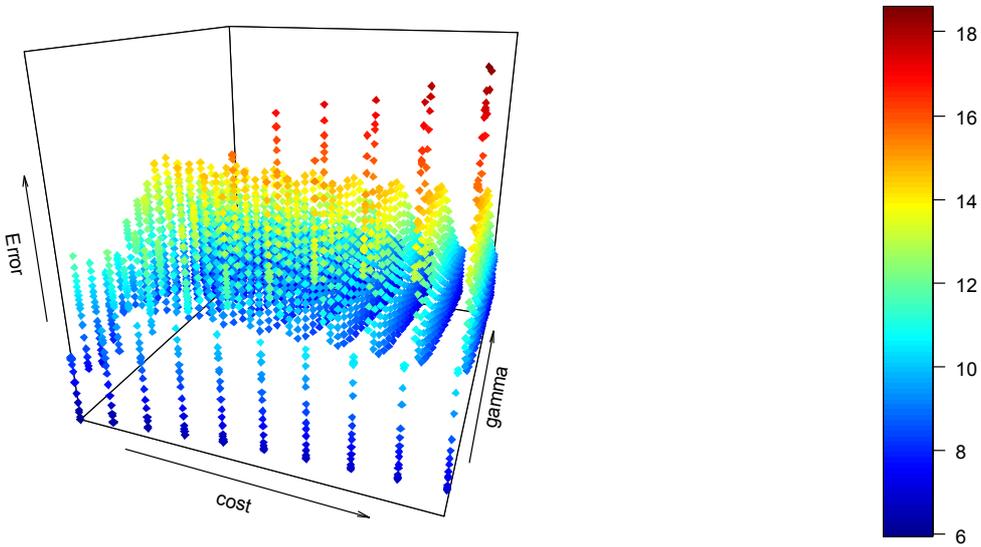


Figure 9. Uganda SVR optimization.

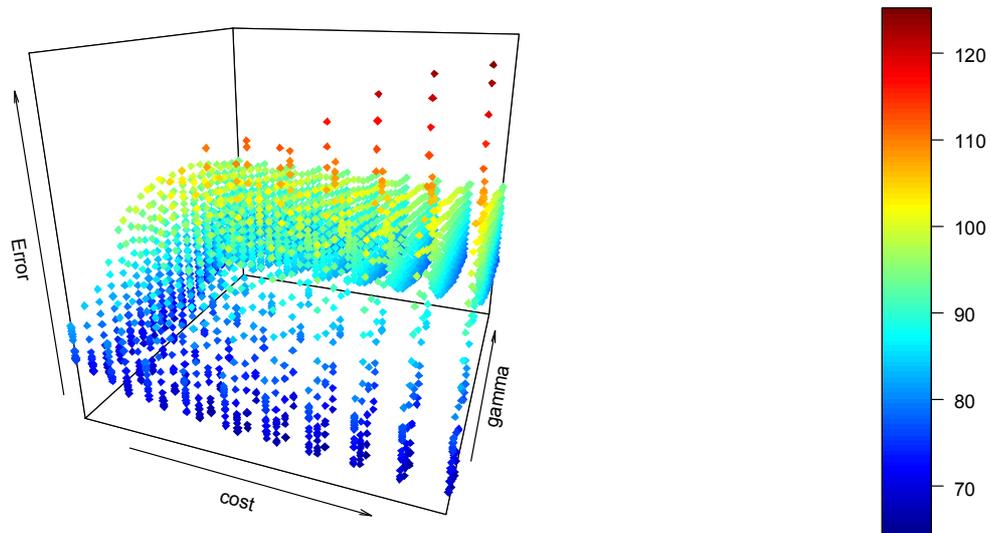


Figure 10. Uganda SVR optimization.

Appendix II: Neural Networks

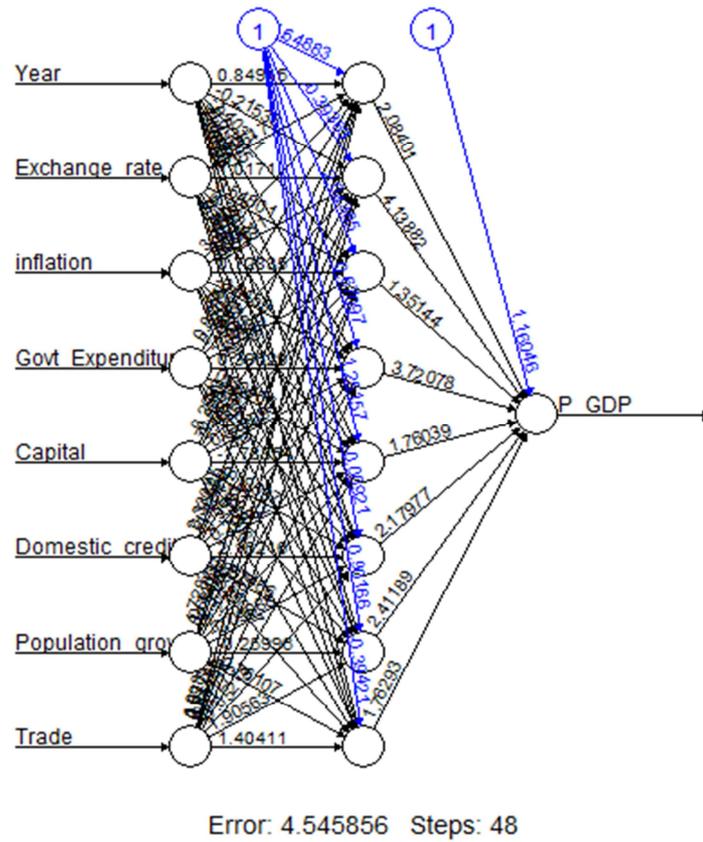


Figure 11. Burundi Artificial Neural Network.

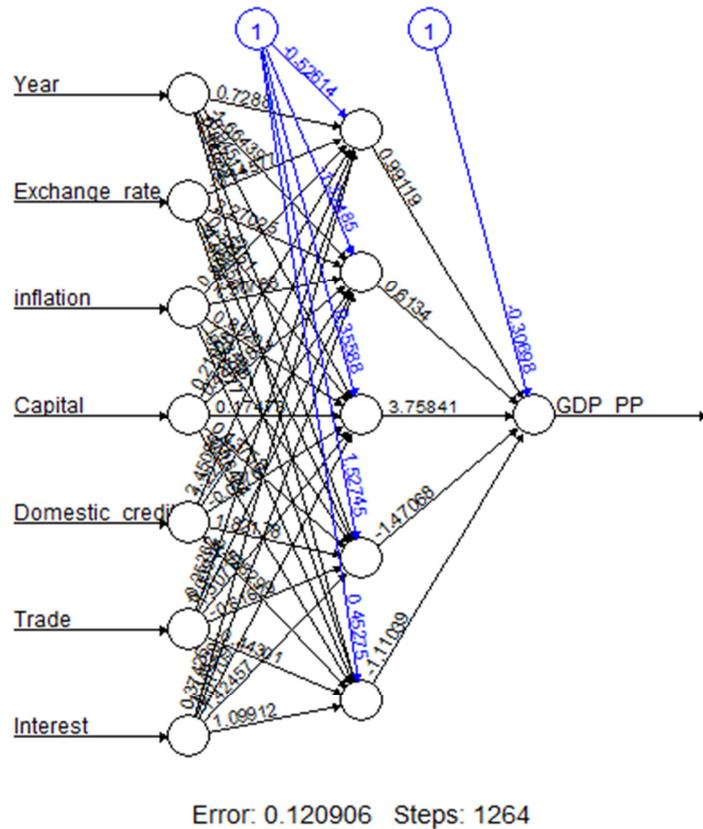
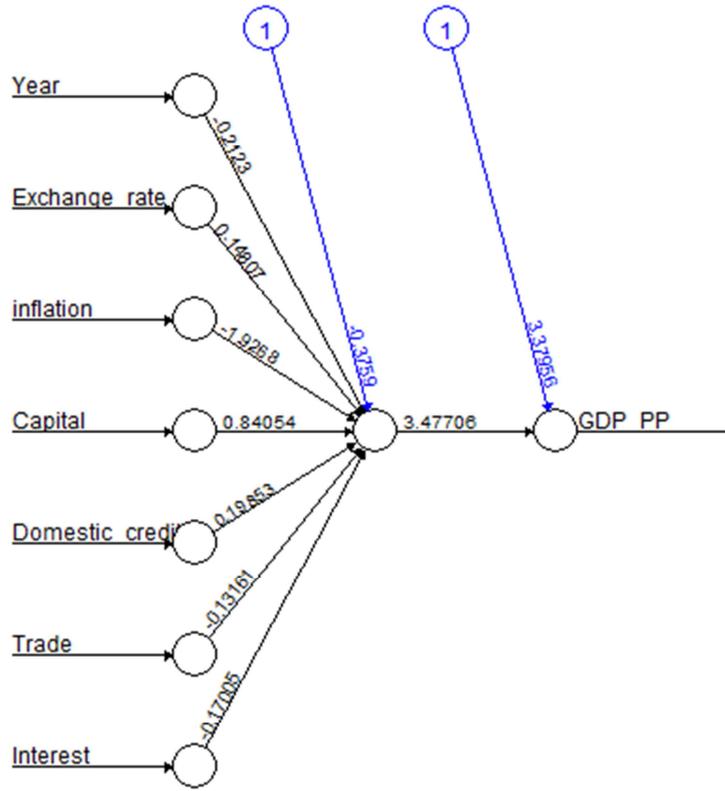
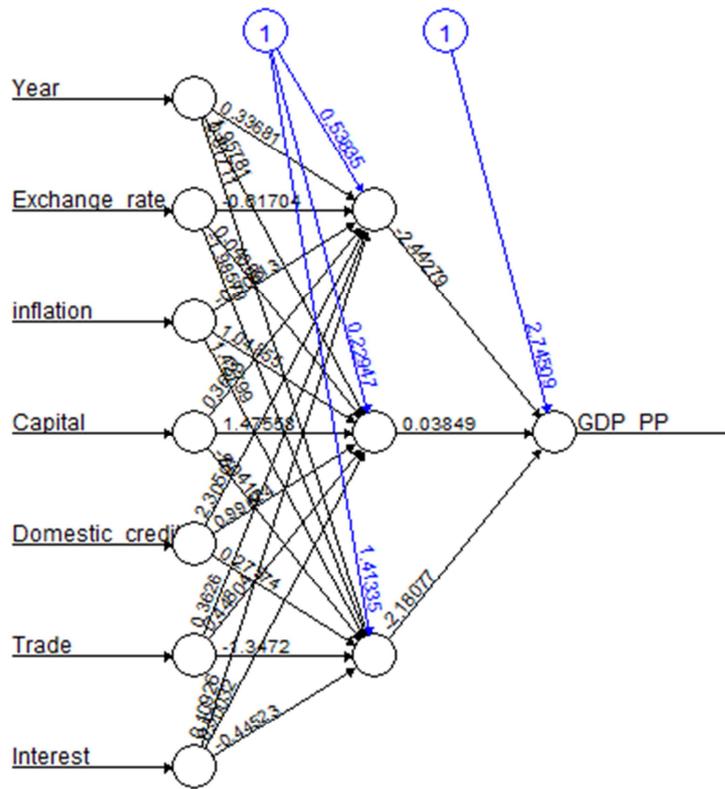


Figure 12. Kenya Artificial Neural Network.



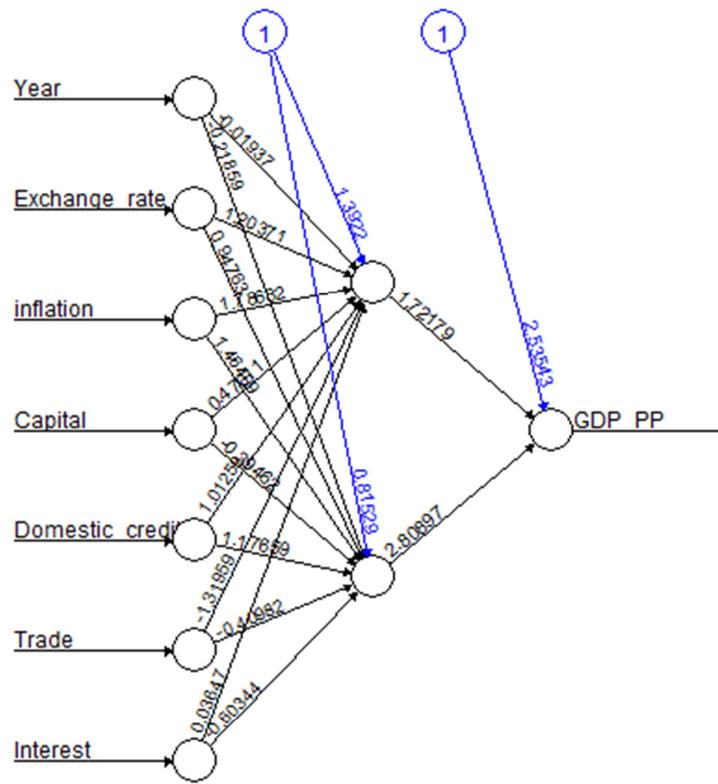
Error: 0.381832 Steps: 3878

Figure 13. Rwanda Artificial Neural Network.



Error: 0.254825 Steps: 1883

Figure 14. Tanzania Artificial Neural Network.



Error: 0.287917 Steps: 2526

Figure 15. Uganda Artificial Neural Network.

Appendix III: Neural Network model

The neural network model is presented in the table 5 below.

Table 5. Weights for neural network models.

Country	Input	Hidden1	Hidden2	Hidden3	Hidden4	Output
Burundi	Bias	0.175	0.692	-0.727	-1.382	0.778
	Year	1.506	1.559	-0.523	0.823	
	Exchange_rate	-0.409	-0.750	-0.235	0.524	
	Government_consumption	-2.460	0.512	0.160	-0.044	
	Gross_fixed_capital	-1.445	0.252	1.395	0.550	
	Trade	-0.395	-0.730	-0.348	2.074	
	Hidden1					-1.124
	Hidden2					0.698
	Hidden3					-1.707
	Hidden4					0.228
Kenya	Bias	-1.340	1.720	0.203		2.284
	Year	-1.029	-0.570	1.918		
	Exchange_rate	0.460	0.416	0.863		
	Government_consumption	-0.952	0.149	0.237		
	Gross_fixed_capital	0.912	-0.283	2.724		
	Trade	0.230	0.910	-0.586		
	Hidden1					0.787
	Hidden2					-1.132
	Hidden3					0.603
	Hidden4					10.862
Rwanda	Bias	-0.582				
	Year	0.202				
	Exchange_rate	-1.711				
	Government_consumption	0.633				
	Gross_fixed_capital	-0.648				
	Trade	0.850				
	Hidden1					-22.339
Tanzania	Bias	-1.337				4.536

Country	Input	Hidden1	Hidden2	Hidden3	Hidden4	Output
Uganda	Year	1.811				
	Exchange_rate	0.853				
	Government_consumption	-0.248				
	Gross_fixed_capital	1.054				
	Trade	0.064				
	Hidden1					0.323
	Bias	0.815				7.021
	Year	-1.558				
	Exchange_rate	-1.179				
	Government_consumption	0.327				
Combined	Gross_fixed_capital	-0.542				
	Trade	-0.762				
	Hidden1					0.025
	Bias	0.947	-0.707	-1.043		4.067
	Country	-1.381	0.744	-0.683		
	Year	-0.130	-0.541	-1.551		
	Exchange_rate	-0.479	-1.437	-1.385		
	Government_consumption	0.918	-0.429	1.699		
	Gross_fixed_capital	-0.798	0.347	-0.182		
	Trade	-0.726	-0.706	-0.916		
	Hidden1				-1.036	
	Hidden2				-0.539	
	Hidden3				0.451	

Acknowledgements

Foremost I acknowledge Almighty God. My supervisors for the guidance through the research work including all the other lecturers in the department for their dedication.

I am highly indebted to my colleague Mr. John M. Ngeny and my brother Vincent Koech for their invaluable advise. Thanks to my friends and classmates for their support.

References

- [1] Swan, T. W. (1956). Economic growth and capital accumulation. *Economic Record*. Wiley. 32 (2): 334–361.
- [2] Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*. 70 (1): 65–94.
- [3] Cobb, C. W. and Douglas, P. H. (1928). A Theory of Production. *American Economic Review*, 139-65.
- [4] Liand, Q. and Pan, C. G. (2005). The Research of High Precision Gray Forecast Model and Its Application in GDP Forecast of 2005. *Research On Financial And Economic Issues*, No. 8, pp. 11-13.
- [5] Kilian, L. and Taylor, M. P. (2003). Why is it So Difficult to Beat the Random Walk Forecast of Exchange Rates?. *Journal of International Economics*, 60, 85-107.
- [6] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2), 357–384.
- [7] Granger, C. W. J. (2001). *Essays in Econometrics: The Collected Papers of Clive W. J. Granger*. Cambridge: Cambridge University Press.
- [8] Diebold, F. X., Nason, J., (1990). Nonparametric exchange rate prediction. *Journal of International Economics* 28 (3-4), 315-332.
- [9] Mizraeh, B. M., (1992). Multivariate nearest-neighbor forecasts of EMS exchange rates. *Journal of Applied Econometrics*, 7, S151–S163.
- [10] Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, 45, 267-290.
- [11] Shin, K., Lee, T., Kim, H., (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, Volume 28, pp. 127-135.
- [12] Molinet, T., J. A. Molinet, M. E. Betancourt, A. Palmer, J. J. Montaña (2015). “Models of Artificial Neural Networks Applied to Demand Forecasting in Nonconsolidated Tourist Destinations”. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. Forthcoming.
- [13] Claveria, O., E. Monte, and S. Torra (2016). “A New Forecasting Approach for the Hospitality Industry”. *International Journal of Contemporary Hospitality Management*, 28 (2).
- [14] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, “Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique,” *Energy and Buildings*, vol. 126, pp. 94–103, 2016.
- [15] W. D. Li, D. M. Kong, and J. R. Wu (2017), “A Novel Hybrid Model Based on Extreme Learning Machine, k-Nearest Neighbor Regression and Wavelet Denoising Applied to Short-Term Electric Load Forecasting”, *Energies*, vol. 10, no. 5, p. 694.
- [16] Shenify, M. et al. (2016). Precipitation Estimation Using Support Vector Machine with Discrete Wavelet Transform. *Water Resource Management*. pp 30: 641.
- [17] Cortes, C.; Vapnik, V. (1995). Support vector networks, In *Proceedings of Machine Learning* 20: 273–297.
- [18] Vapnik, V. N. (1998). *Statistical Learning Theory*. New York: John Wiley and Sons.

- [19] Peng, Y.; Kou, G.; Shi, Y.; Chen, Z. X. (2008). A descriptive framework for the field of data mining and knowledge discovery. *International Journal of Information Technology and Decision Making*, 7 (4): 639–682.
- [20] Yang, Q.; Wu, X. D. (2006). 10 Challenging problems in data mining research. *International Journal of Information Technology and Decision Making* 5 (4): 567–604.
- [21] Cassel, G. (1918). Abnormal Deviations in International Exchanges. *The Economic Journal*, 28, No. 112 (112): 413–415.
- [22] Bassanini, A., Scarpetta, A. and Hemmings, P. (2011), *Economic Growth: The Role of Policies and Institutions. Panel Data Evidence from OECD Countries* (Working Paper No. 283). OECD Economics Department.
- [23] Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints. Master's thesis, Department of Mathematics, Univ. of Chicago.
- [24] Kuhn H. W. and Tucker A. W. (1951). Nonlinear programming. In: *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probabilistics*, Berkeley. University of California Press, pp. 481–492.
- [25] Serneels, P. and Verpoorten, M (2012). The Impact of Armed Conflict on Economic Performance: Evidence from Rwanda. Discussion Paper No. 6737. IZA.
- [26] Lopez, H., Wodon, Q. (2005). The Economic Impact of Armed Conflict in Rwanda. *Journal of African Economies*, 14 (4): 586-602.
- [27] Demiriz, A., Bennett, K., Breneman, C. and Embrechts, M. (2001). *Support vector machine regression in chemometrics*. Computing Science and Statistics, 2001.