

Modelling and Forecasting Unemployment Trends in Kenya Using Advanced Machine Learning Techniques

Reuben Opondo*, Davis Bundi, Patrick Weke

Department of Mathematics, University of Nairobi, Nairobi, Kenya

Email address:

opondoreuben94@gmail.com (Reuben Opondo), dbundi@uonbi.ac.ke (Davis Bundi), pweke@uonbi.ac.ke (Patrick Weke)

*Corresponding author

To cite this article:

Reuben Opondo, Davis Bundi, Patrick Weke. (2024). Modelling and Forecasting Unemployment Trends in Kenya Using Advanced Machine Learning Techniques. *American Journal of Theoretical and Applied Statistics*, 13(6), 242-254.

<https://doi.org/10.11648/j.ajtas.20241306.15>

Received: 21 October 2024; **Accepted:** 18 November 2024; **Published:** 18 December, 2024

Abstract: This study conducts a comprehensive analysis exploring the relationship between key macroeconomic indicators and the unemployment rate, alongside evaluating the predictive accuracy of modern regression models. The correlation analysis examines the association between unemployment rate percentages and five macroeconomic variables: real Gross Domestic Product (GDP) growth, gross public debt as a percentage of GDP, population size, government revenue as a percentage of GDP, and government expenditure as a percentage of GDP. The results highlight significant correlations, particularly the strong positive relationship between unemployment rates and gross public debt (% of GDP) (0.8417), while real GDP growth shows a weak correlation (0.0783), indicating that debt levels may be a more crucial determinant of unemployment variations in this context. Additionally, a comparison of modern regression models, namely Support Vector Regression (SVR), Neural Network Regression, and Bayesian Regression, is conducted based on their performance metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R-Squared, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Among the models, Support Vector Regression outperforms the others, with the lowest MAE (0.0823), RMSE (0.0878), and the highest R-Squared value (0.9915), along with notably favourable AIC (-100.2072) and BIC (-69.4268) scores. Neural Network Regression also delivers competitive performance with a slightly higher MAE and RMSE but a similarly strong R-Squared (0.9887). In contrast, Bayesian Regression exhibits weaker predictive power with higher error metrics (MAE = 0.2579, RMSE = 0.3109) and a significantly lower R-Squared (0.8806), AIC (28.0408), and BIC (36.8352). These findings underscore the efficacy of SVR in predictive modelling for macroeconomic datasets, suggesting its suitability for unemployment rate forecasting.

Keywords: Modelling, Machine Learning, Kenya, Unemployment Trends

1. Introduction

Unemployment rate is the share of individuals in the workforce who are jobless. To date, the biggest hurdle facing Kenyan economic growth has been the high unemployment rate caused by political instability and high rates of corruption. Among the youth in the country a large proportion are encountering joblessness as compared to the rest of the population. The unemployment rate of the youth was approximately 25% in the year 2005/6 while the general unemployment rate in the country was 12.7% which is half of the youth. There was a reduction of the unemployment

rate by 15.8% 13.1% among the age of 15-19 and 20-24 respectively. The total unemployment rate reduced by 8.6%. Between the rural and urban areas, the unemployment rate varies significantly. The youth unemployment seemed to dominate unlike that of the working-age population. For instance, in 2009, the overall unemployment rate was about 11.0% in the urban areas relative to youth unemployment rate of 19.1% [11].

From the mid-1960s to the early 1980s, there was an inflation imbalance in the United States. Consumer Price Index (CPI) is termed as inflation rate by the World Bank.

Hence, inflation refers to the increase of cost goods and services being acquired from economy by consumers on a yearly basis. Hence, the inflation rate measures changes over a certain period to be that of purchasers' costs or Gross Domestic Product (GDP) deflator incorporating pricing goods and services produced domestically [8]. CPI is actually measures how the general population in this nation might need to spend for day-to-day living Inflation as a phenomenon can be generalised with price levels instead of individual items, like the prices for goods and services. There is a wide range of anomalies in economies from wealthy to those more impoverished as there are with states and their key regions [3].

The targets for economic growth within Kenya Vision 2030's medium-term plan have not been met. For example, the economy grew at 6 percent in 2006 and was targeted to grow by a sustainable rate of about 10% by around 2012/2013. Results from studies on the linkage between unemployment rate and economic growth are mixed. Also in their study on youth unemployment-turnover nexus with economic growth evidence from Central & Eastern European countries found that so long as the problem seemed to be a great concern of significance exclusive economic growth would fail to solve its reduction concluding that there is no significant direct impact between poverty alleviation and income inequality [7]. A survey in the US economy showed that as GDP increases, unemployment rate decreases. But with the GDP decelerating, unemployment was already in free fall [4]. To be slightly more nuanced it was found that a return to long run growth causes employment and living standards to rise hence in the longer term unemployment should fall [13].

Okun's law indicates that a one percent surge in unemployment is due to a negative growth of two percent in GDP. A study on Kenya found that the coefficient of the two variables to be 0.12 instead of 0.3 as indicated in the Okun's law [10].

The Solow growth model predicts that an economy will grow due to the increase in production per worker induced by raising a sustained level of investment but at diminishing rates affected from increased levels of capital accumulation relative to labor. The marginal product of each additional unit of capital falls over time gradually bringing us back on to a long-run growth path where real GDP expands along side workforce expansion & improvements in productivity. The baseline growth path makes output, capital and labor grow at the same constant rates - so preserving a constant ratio of out-turn per employee; resource per employee. Neoclassical economists implicitly assume more labour and capital can only be employed if their productivity improves as this is the object of general increase in efficiency writing to improve overall growth, he argue higher employment will result from a pro-growth monetary policy combining increased output given demand levels with lean manufacturing or other measures raising factor/productivity. Differences in technology play a huge role and represent the vast differences between national growth rates. Add to this the concepts of 'catch-up growth' that lie at the heart of the Solow model itself, whereby less developed countries grow faster than richer ones can because

they have higher marginal returns on investment. While the model predicts living standard convergence, as measured by per capita income, how much this catch-up can erase disparities in standards of living is open to dispute. As such, the model highlights capital accumulation volume of labor and technological change as main determinants in the functioning of an economy [12].

As much in his study, by giving expression to the Quantity Theory of Money (QTM), which posits that the general cost level for goods & services is directly linked with money supply within an economy. That means, in simple terms that if the amount of money doubles all other things equal then so will prices and you would end up having to pay twice as much for roughly same thing. What this relationship tells us is a direct connection between the level of prices or, in other words inflation and it mainly defines measures changes within economies how fast either goods or services cost to increase [5]. Also evaluated the QTM, finding support for a connection between money supply and price level [1]. However, the most classic notion of this theory has been viewed with skepticism from modern economists. In the same way that other commodities are determined by supply and demand forces also do so with money [2]. When money gains a lot, its marginal value also decreases and buying power appears to be diluted. The amount of money moving in the economy has many robust impacts across economic health [9]. Small increases in the money supply can have a significant impact on both price levels and market goods/services. This is what is confirmed as the the intervention of money supply in different areas of economy. The setting of targets for money growth is favoured coupled with encouragement to allow such measures being put in place so as not choke economic development and recession [14]. Considering that the earlier studies focused on the youth unemployment, this study will be focusing specifically on overall unemployment rate and determine if the key coefficients still stand, therefore, examine the causal relationship between unemployment and other key indicators including economic growth.

2. Methodology

2.1. Support Vector Regression (SVR)

At its core, the idea behind Support Vector Regression (SVR) is to keep the error of one data point within a certain epsilon range and some threshold called epsilon-insensitive tube. SVR performs well in situations where we need to penalise the outliers more heavily than their fair contributions or when the data is too extreme. This is particularly precious for financial time series, which is a perfect representative of volatility and noisy data or any system where the stability of forecasting against outliers has greater importance. The capability of SVR to handle small, acceptable error and ignore them so focusing on a larger deviations makes the model useful in real-world situations where changes are unpredictable [6].

The formula for SVR can be represented in its basic form as

follows:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad f(x) = \langle w, x \rangle + b \quad (2)$$

leading to

$$\begin{aligned} y_i - w \cdot x_i - b &\leq \epsilon + \xi_i, \\ w \cdot x_i + b - y_i &\leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* &\geq 0, \end{aligned} \quad \phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i^- + \xi_i^+) \quad (3)$$

for all i .

where:

1. w Is the vector of weight.
2. b stands for bias term.
3. ξ_i and ξ_i^* stand for additional variables measuring the degree of violation of the epsilon-insensitive tube for data points that are above or below the tube, respectively.
4. C parameter for regularisation.
5. ϵ represents the width of the epsilon-insensitive tube and controls model's sensitivity to deviations from the actual values within this range.

Derivation of support vector machine regression

Let new data denoted as

$$\mathcal{D} = \{(x_1, y_1), \dots, (x_l, y_l)\}, \quad x \in \mathbb{R}^d, \quad y \in \mathbb{R} \quad (1)$$

Linear function is denoted by

We can let an optimal function in a regression be

With $C > 0$ is an already denoted constant value, determining the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated, and ξ are slack variables representing upper ξ^+ and lower ξ^- constraints on the outputs of the system.

ϵ - insensitive Loss Function

The ϵ -insensitive loss function is denoted by

$$L(y, f(x)) = \max(0, |y - f(x)| - \epsilon) \quad (4)$$

This loss function only penalizes deviations of the prediction from the actual values that are greater than ϵ , hence promoting a model that is insensitive to smaller errors.

This loss function is ideal when small amounts of error are acceptable, which is also referred to as a 'soft-margin'. In ϵ -insensitive loss function, any points within some selected range ϵ are considered to have no error at all, which means the ϵ -insensitive loss function can be represented as

$$L_\epsilon(y) = \begin{cases} 0 & \text{for } |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon & \text{otherwise} \end{cases} \quad (5)$$

This error-free margin makes the loss function an ideal candidate for support vector regression.

In linear function $f(x) = \langle w, x \rangle + b$, $\langle w, x \rangle$ denotes the dot product in \mathbb{R}^d . Since we mentioned before, one of our goals is to ensure the flatness, which means seeking a small w . One of our ways is to minimize the norm, i.e., $\|w\|^2 = \langle w, w \rangle$.

Minimize:

$$\frac{1}{2} \|w\|^2 \quad (6)$$

Subject to:

$$y_i - \langle w, x_i \rangle - b \leq \epsilon \quad (7)$$

$$\langle w, x_i \rangle + b - y_i \leq \epsilon \quad (8)$$

Applying ϵ -insensitive loss function for minimizing the optimal regression function,

Minimize:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i^- + \xi_i^+) \quad (9)$$

Subject to:

$$\begin{aligned} y_i - \langle w, x_i \rangle - b &\leq \epsilon + \xi_i^- \\ \langle w, x_i \rangle + b - y_i &\leq \epsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ &\geq 0 \end{aligned} \quad (10)$$

Here, the key idea is to construct a Lagrange function [15]. We proceed as follows:

$$\begin{aligned} L(\eta_i, \eta_i^*, \alpha_i, \alpha_i^*) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i^- + \xi_i^+) - \sum_{i=1}^l (\eta_i \xi_i^- + \eta_i^* \xi_i^+) - \sum_{i=1}^l \alpha_i (\epsilon + \xi_i^- - y_i + \langle w, x_i \rangle + b) \\ &\quad - \sum_{i=1}^l \alpha_i^* (\epsilon + \xi_i^+ - y_i + \langle w, x_i \rangle + b) \end{aligned} \quad (11)$$

Here L is the Lagrangian and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrange multipliers. Taking partial derivatives of Equation (11) with primal variables (w, b, ξ_i^-, ξ_i^+) :

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (12)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \quad (13)$$

$$\frac{\partial L}{\partial \xi_i^+} = C - \alpha_i^* - \eta_i^* = 0 \quad (14)$$

Substituting the derivatives from Equations (12) to (14) into L , the solution is given by,

$$\max -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle + \sum_{i=1}^l \alpha_i (y_i - \epsilon) - \alpha_i^* (y_i + \epsilon) \quad (15)$$

Alternatively,

$$\min \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i + \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon \quad (16)$$

with constraints.

$$0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, l \quad \min \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \beta_i \beta_j \langle x_i, x_j \rangle - \sum_{i=1}^l \beta_i y_i \quad (23)$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (17) \quad \text{with constraints,}$$

$$-C \leq \beta_i \leq C, \quad i = 1, \dots, l \quad (24)$$

Equation (13) can be rewritten as

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (18) \quad \sum_{i=1}^l \beta_i = 0 \quad (25)$$

Thus, the function $f(x)$ can be expressed as

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (19) \quad w = \sum_{i=1}^l \beta_i x_i \quad (26)$$

This is called Support Vector expansion, where w can be described as a linear combination of x_i .

The Karush-Kuhn-Tucker (kkt) conditions that are satisfied by the solution are,

$$\alpha_i \alpha_i^* = 0, \quad i = 1, \dots, l \quad (20)$$

which is also called complementary slackness [15]. This condition shows that there can never be a set of dual variables which are both α_i and α_i^* simultaneously nonzero, which allows us to conclude that

$$\epsilon - y_i + \langle w, x_i \rangle + b \geq 0 \text{ and } \xi_i = 0 \text{ if } \alpha_i < C \quad (21)$$

$$\epsilon - y_i + \langle w, x_i \rangle + b \leq 0 \text{ if } \alpha_i = 0 \quad (22)$$

Therefore, the support vectors said to be points in places exactly one of the Lagrange multipliers is beyond zero. When $\epsilon = 0$, we therefore obtain L loss function making the optimisation problem simplified

and the regression function is given by Equation (21), where

$$b = -\frac{1}{2} \langle w, (x_r + x_s) \rangle \quad (27)$$

where x_r and x_s are support vectors lying on the edge of the margin.

In addition, if $\epsilon = 0$, the result is just the median regression.

2.2. Neural Network Regression

Neural Network Regression is used to predict continuous output values using a neural network that can capture Non-linear and complex relationships between inputs features with the target variable which Linear regression models are not capable of. This is a type of linear regression via an artificial neural network with - input layer having data features, multiple hidden layers that processes these features through weighted sum and series non-linear activation function followed by the output layer gives as continuous prediction.

The way the network 'learns' is through training, during which it tries to minimise a given loss function (typically mean squared error between what it predicts and the ground truth) It does this by iteratively optimising the network's internal

parameters (weights + biases) via an optimisation approach like gradient descent in order to minimise predictive errors. Let the predicted output \hat{y} be expressed as:

$$\hat{y} = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

Where:

1. x_i predictor variables
2. w_i weight per input feature.
3. b stands for bias term, which helps the model make

better fits.

4. $\sum_{i=1}^n w_i x_i + b$ represents the weighted sum of the inputs plus the bias.
5. f activation function, which is important in introducing non-linearity into the model (e.g., ReLU, Sigmoid).

Derivation of neural network regression - Gradient of the Weights

To define our “outer function”, we start again in layer l and consider the loss function to be a function of the weighted inputs z :

$$\mathcal{L}(z_1^{[l]}, z_2^{[l]}, \dots, z_{n^{[l]}}^{[l]}) \quad (28)$$

To define our “inner functions”, we take again a look at the forward propagation equation:

$$z_k^{[l]} = w_{k,1}^{[l]} a_1^{[l-1]} + w_{k,2}^{[l]} a_2^{[l-1]} + \dots + w_{k,n^{[l-1]}}^{[l]} a_{n^{[l-1]}}^{[l-1]} + b_k^{[l]}, \quad \forall k \in \{1, \dots, n^{[l]}\} \quad (29)$$

and notice, that z is a function of the elements of weight matrix W :

$$z_k^{[l]} = z^{[l]} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right) \quad (30)$$

The resulting nested function depends on the elements of W :

$$\mathcal{L} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right) = \mathcal{L} \left(z_1^{[l]} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right), \dots, z_{n^{[l]}}^{[l]} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right) \right) \quad (31)$$

We apply the chain rule:

$$\sum_{k=1}^{n^{[l]}} \frac{\partial \mathcal{L}}{\partial z_k^{[l]}} \left(z_1^{[l]} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right), \dots, z_{n^{[l]}}^{[l]} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n^{[l-1]},n^{[l-1]}}^{[l]} \right) \right) \frac{\partial z_k^{[l]}}{\partial W_{i,j}^{[l]}} = \quad (32)$$

As before the first term in the above expression is the error of layer l and the second term can be evaluated to be:

$$\frac{\partial z_k^{[l]}}{\partial W_{i,j}^{[l]}} = \begin{cases} a_j^{[l-1]} & \text{if } k = i, \\ 0 & \text{otherwise.} \end{cases} \quad (33)$$

as we will quickly show. To this end, we first notice that each weighted input z depends only on a single row of the weight matrix W :

$$\begin{bmatrix} z_1^{[l]} \\ z_2^{[l]} \\ \vdots \\ z_{n^{[l]}}^{[l]} \end{bmatrix} = W^{[l]} \begin{bmatrix} a_1^{[l-1]} \\ a_2^{[l-1]} \\ \vdots \\ a_{n^{[l-1]}}^{[l-1]} \end{bmatrix} + b^{[l]} = \begin{bmatrix} W_{1,1}^{[l]} & W_{1,2}^{[l]} & \dots & W_{1,n^{[l-1]}}^{[l]} \\ W_{2,1}^{[l]} & W_{2,2}^{[l]} & \dots & W_{2,n^{[l-1]}}^{[l]} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n^{[l]},1}^{[l]} & W_{n^{[l]},2}^{[l]} & \dots & W_{n^{[l]},n^{[l-1]}}^{[l]} \end{bmatrix} \begin{bmatrix} a_1^{[l-1]} \\ a_2^{[l-1]} \\ \vdots \\ a_{n^{[l-1]}}^{[l-1]} \end{bmatrix} + \begin{bmatrix} b_1^{[l]} \\ b_2^{[l]} \\ \vdots \\ b_{n^{[l]}}^{[l]} \end{bmatrix} \quad (34)$$

Hence, taking the derivative with respect to coefficients from other rows, must yield zero:

$$\frac{\partial z_2^{[l]}}{\partial W_{3,2}^{[l]}} = \frac{\partial z_2^{[l]}}{\partial W_{1,1}^{[l]}} \cdot a_1^{[l-1]} + \frac{\partial z_2^{[l]}}{\partial W_{2,2}^{[l]}} \cdot a_2^{[l-1]} + \dots + \frac{\partial z_2^{[l]}}{\partial W_{n^{[l]},n^{[l-1]}}^{[l]}} \cdot a_{n^{[l-1]}}^{[l-1]} + \frac{\partial z_2^{[l]}}{\partial b_{3,2}^{[l]}} = 0 \quad (35)$$

In contrast, when we take the derivative with respect to elements of the same row, we get:

$$\frac{\partial z_2^{[l]}}{\partial W_{2,2}^{[l]}} = \frac{\partial z_2^{[l]}}{\partial W_{2,1}^{[l]}} \cdot a_1^{[l-1]} + \frac{\partial z_2^{[l]}}{\partial W_{2,2}^{[l]}} \cdot a_2^{[l-1]} + \dots + \frac{\partial z_2^{[l]}}{\partial W_{2,n^{[l-1]}}^{[l]}} \cdot a_{n^{[l-1]}}^{[l-1]} + \frac{\partial z_2^{[l]}}{\partial b_2^{[l]}} = a_2^{[l-1]} \quad (36)$$

Thus we arrive at the final formula:

$$\frac{\partial \mathcal{C}}{\partial W_{i,j}^{[l]}} \left(W_{1,1}^{[l]}, W_{1,2}^{[l]}, \dots, W_{n,n^{[l-1]}}^{[l]} \right) = \sum_{k=1}^{n^{[l]}} \frac{\partial \mathcal{C}}{\partial z_k^{[l]}} \frac{\partial z_k^{[l]}}{\partial W_{i,j}^{[l]}} = \frac{\partial \mathcal{C}}{\partial a_i^{[l-1]}} \cdot a_i^{[l-1]} \quad (37)$$

Expressing the formula in matrix form for all values of i and j gives us:

$$\frac{\partial \mathcal{L}}{\partial W^{[l]}} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial z_1^{[l]}} a_1^{[l-1]} & \frac{\partial \mathcal{L}}{\partial z_1^{[l]}} a_2^{[l-1]} & \dots & \frac{\partial \mathcal{L}}{\partial z_1^{[l]}} a_{n^{[l-1]}}^{[l-1]} \\ \frac{\partial \mathcal{L}}{\partial z_2^{[l]}} a_1^{[l-1]} & \frac{\partial \mathcal{L}}{\partial z_2^{[l]}} a_2^{[l-1]} & \dots & \frac{\partial \mathcal{L}}{\partial z_2^{[l]}} a_{n^{[l-1]}}^{[l-1]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathcal{L}}{\partial z_{n^{[l]}}^{[l]}} a_1^{[l-1]} & \frac{\partial \mathcal{L}}{\partial z_{n^{[l]}}^{[l]}} a_2^{[l-1]} & \dots & \frac{\partial \mathcal{L}}{\partial z_{n^{[l]}}^{[l]}} a_{n^{[l-1]}}^{[l-1]} \end{bmatrix} \in \mathbb{R}^{n^{[l]} \times n^{[l-1]}} \quad (38)$$

and can compactly be shown using the following familiar outer product:

$$\frac{\partial \mathcal{L}}{\partial W^{[l]}} = \frac{\partial \mathcal{L}}{\partial z^{[l]}} \cdot (a^{[l-1]})^T \quad (39)$$

with

$$\frac{\partial \mathcal{L}}{\partial W^{[l]}} \in \mathbb{R}^{n^{[l]} \times n^{[l-1]}}, \quad \frac{\partial \mathcal{L}}{\partial z^{[l]}} \in \mathbb{R}^{n^{[l]}}, \quad a^{[l-1]} \in \mathbb{R}^{n^{[l-1]}} \quad (40)$$

2.3. Bayesian Regression

Bayesian Regression is a linear model that performs the statistical problem entirely in terms of probability. Bayesian regression allows for the last property by modeling coefficients as random variables with prior distributions (as opposed to traditional methods which produce fixed coefficient estimates). This prior information might correspond to either well-established priors reflecting existing beliefs or knowledge about the parameters in question, before observing any data.

At the core of Bayesian regression is to apply Bayes' theorem in order to update prior beliefs on coefficients given observed data resulting into posterior distribution over these coefficients. Not only the parameters get estimates, but we

also capture this in terms of certainties for each estimate using posterior distributions.

Such a probabilistic framework is especially beneficial when we have limited or missing data, and controlling the uncertainty plays an important role. Incorporating this prior knowledge, and constantly updating it with the incoming data is what enables more nuancedness than regular linear regression. As such, it is a very versatile tool for predictive analysis and particularly useful in areas where decisions are to be taken under uncertainty.

Posterior of the bayesian regression model

The likelihood, prior on σ^2 and conditional prior on β , are given by:

$$\begin{aligned} p(y | X, \beta, \sigma^2) &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left(-\frac{1}{2\sigma^2} (y - X\beta)^\top (y - X\beta) \right), \\ p(\beta | \sigma^2) &= (2\pi\sigma^2)^{-\frac{P}{2}} |\Lambda_0|^{\frac{1}{2}} \exp \left(-\frac{1}{2\sigma^2} (\beta - \mu_0)^\top \Lambda_0 (\beta - \mu_0) \right), \\ p(\sigma^2) &= \frac{b_0^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-(a_0+1)} \exp \left(-\frac{b_0}{\sigma^2} \right). \end{aligned} \quad (41)$$

Likelihood's Gaussian kernel can be decomposed and combined with the prior's kernel as shown below

$$\begin{aligned} (y - X\beta)^\top (y - X\beta) + (\beta - \mu_0)^\top \Lambda_0 (\beta - \mu_0) &= (y - X\hat{\beta})^\top (y - X\hat{\beta}) + (\hat{\beta} - \beta)^\top X^\top X (\hat{\beta} - \beta) \\ &\quad + (\beta - \mu_0)^\top \Lambda_0 (\beta - \mu_0) \\ &= y^\top y + \mu_0^\top \Lambda_0 \mu_0 - \mu_N^\top \Lambda_N \mu_N + (\beta - \mu_N)^\top \Lambda_N (\beta - \mu_N) \end{aligned} \quad (42)$$

Where μ_N and Λ_N are defined as:

$$\Lambda_N = X^\top X + \Lambda_0, \quad (43)$$

$$\mu_N = \Lambda_N^{-1} (\Lambda_0 \mu_0 + X^\top y). \quad (44)$$

Rewriting posterior

$$p(y | X, \beta, \sigma^2) \propto (2\pi\sigma^2)^{-\frac{P}{2}} |\Lambda_0|^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\beta - \mu_N)^\top \Lambda_N(\beta - \mu_N)\right) \\ \times (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2}[y^\top y + \mu_0^\top \Lambda_0 \mu_0 - \mu_N^\top \Lambda_N \mu_N]\right) \times \frac{b_0^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-(a_0+1)} \exp\left(-\frac{b_0}{\sigma^2}\right). \quad (45)$$

Results to

$$(\sigma^2)^{-(a_0+\frac{N}{2}+1)} \exp\left(-\frac{1}{\sigma^2} \left[b_0 + \frac{1}{2} \{y^\top y + \mu_0^\top \Lambda_0 \mu_0 - \mu_N^\top \Lambda_N \mu_N\}\right]\right). \quad (46)$$

Where:

$$a_N = a_0 + \frac{N}{2}, \quad (47)$$

$$b_N = b_0 + \frac{1}{2} (y^\top y + \mu_0^\top \Lambda_0 \mu_0 - \mu_N^\top \Lambda_N \mu_N). \quad (48)$$

Posterior summary becomes

$$p(\beta, \sigma^2 | X, y) \propto p(\beta | X, y, \sigma^2) p(\sigma^2 | X, y) \quad (49)$$

$$(50)$$

where

$$\beta | X, y, \sigma^2 \sim \mathcal{N}_P(\mu_N, \sigma^2 \Lambda_N^{-1}), \quad (51)$$

$$\sigma^2 | y, X \sim \text{InvGamma}(a_N, b_N). \quad (52)$$

2.4. Data Source

This study used secondary time series data. The secondary data was obtained from the Development Indicators of the International Monetary Fund (IMF). The data is annual data ranging from 1991 to 2023.

3. Results and Discussion

This study modelled unemployment rate in Kenya (dependent variable) using Real GDP growth, Gross public

debt, % of GDP, Population, Government revenue, % of GDP and Government expenditure, % of GDP as independent variables. Therefore, the data was collected and presented in the following Table 1. Economic Data from 1991 to 2023 (see in the appendix).

Table 1. Kenya's Economic Data from 1991 to 2023.

| Year | ur | growth | gpb | pop | grp | gep |
|------|------|--------|-------|-------|-------|-------|
| 1991 | 2.86 | 1.30 | 43.04 | 22.70 | 9.88 | 16.44 |
| 1992 | 3.03 | -1.10 | 41.16 | 23.40 | 9.33 | 17.60 |
| 1993 | 3.16 | -0.10 | 61.64 | 24.10 | 10.35 | 18.92 |
| 1994 | 3.12 | 2.50 | 57.01 | 24.80 | 14.65 | 18.80 |
| 1995 | 2.95 | 4.30 | 52.08 | 25.50 | 17.10 | 17.33 |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |
| 2023 | 5.60 | 5.50 | NA | 51.54 | NA | NA |

3.1. Descriptive Statistics

Descriptive statistics are simply different ways to describe the data. Descriptive analysis includes few key statistics such as: mean, median, standard deviation, variance and range.

Table 2. Statistical Summary of Various Indicators.

| Indicator | Unit | Mean | Median | SD | Min | Max | Range | Kurtosis | Skew |
|---------------------------------|--------------------|-------|--------|-------|-------|-------|-------|----------|-------|
| Unemployment rate | Annual % change | 3.33 | 2.93 | 0.99 | 2.62 | 5.69 | 3.07 | 0.95 | 1.62 |
| Real GDP growth | Annual % change | 3.62 | 4.20 | 2.38 | -1.10 | 8.10 | 9.20 | -0.84 | -0.29 |
| Gross public debt % of GDP | % of GDP | 45.47 | 41.28 | 10.49 | 34.19 | 67.94 | 33.74 | -0.54 | 0.90 |
| Population | Millions of people | 36.20 | 35.70 | 8.86 | 22.70 | 51.54 | 28.84 | -1.31 | 0.14 |
| Government revenue % of GDP | % of GDP | 15.75 | 16.74 | 2.26 | 9.33 | 17.98 | 8.66 | 1.66 | -1.52 |
| Government expenditure % of GDP | % of GDP | 19.41 | 18.86 | 3.81 | 13.68 | 25.34 | 11.66 | -1.54 | 0.18 |

An annual percent change in unemployment is expressed as the rate, with an average of 3.3339% and a median of 2.9300%. The unemployment Standard Deviation (SD) = 0.9874% suggesting moderate variability This dataset has a normal range from 2.6200%, the lowest value to highest being 5.6900% causing Interval of total length as:3.0700%. The kurtosis of 0.9499 indicated that the distribution is slightly

more peaked than a Gaussian process. A skew of 1.6175 means that unemployment rates tend to be skewed high.

Average annual growth for real GDP, which is the same as the rate of change (expressed in % per annum), has a mean value of 3.6152% and median value much higher at 4.2000%. The SD of 2.3803%, this lays the ground for significant global variations in terms of national rates. It ranges

from -1.1000% (indicating there are times of contraction in the economy) to a maximum value of 8.1000%, and has a range of 9.2000%. Kurtosis and Skew-ness were -0.8429 (suggesting thinner distribution than normal) and -0.2937 (suggesting a slight left skew: towards more no of data points in the higher side leading to an increase for median compared mean).

Mean of the gross public debt as percentage of GDP is 45.4685% and median equals to: 41.2783%. SD of 10.4890% reveals a medium level of variability. The debt percentages vary from the lowest at 34.1945% to a maximum of 67.9362%, giving us an interval of 33.7417%. Its negative (kurtosis of -0.5425) skew makes it a bit flatter than the normal distribution, and its positive (skewness of 0.9038) asymmetry skews the tail to the right as higher values have more influence on mean vs median;

Population (millions of people; mean = 36.2037, median = 35.7000) The substantial SD of 8.8597 million lends itself to a high variance in the size and scale of populations. We have a low of 22.7000 million, and a high of 51.5390 million

contributing to the range value which is calculated as $\text{High} - \text{Low} = \text{Range}$ (28.8390).

Government Revenue (% of GDP): Mean: 15.7508%, Median: 16.7396% (% of the GDP). Its 2.2621% SD points to moderate variability, Its minimum recorded value is 9.3252% and its maximum: 17.9845%, with a range of only 8.6593%. The positive kurtosis of 1.6552 means a taller distribution than the normal one, while negative skewness of -1.5245 tells us that again: everything by default with more values lower under mean then median value (to right).

Government Expenditure (% of GDP): The mean is 19.4147% and the median 18.8558% (as a percent of GDP). The SD is 3.8100%, indicating much fluctuation in expenditure percentages. It varies from 13.6822 to its highest value, which is the 25.3444% and with a range of this variable equals an amount such as 11.6622%. Kurtosis is negative at -1.5354, indicating a flat distribution and Skewness of 0.1797 showing slight right skewed distribution.

The unemployment rate trend in Kenya from 1991 to 2023

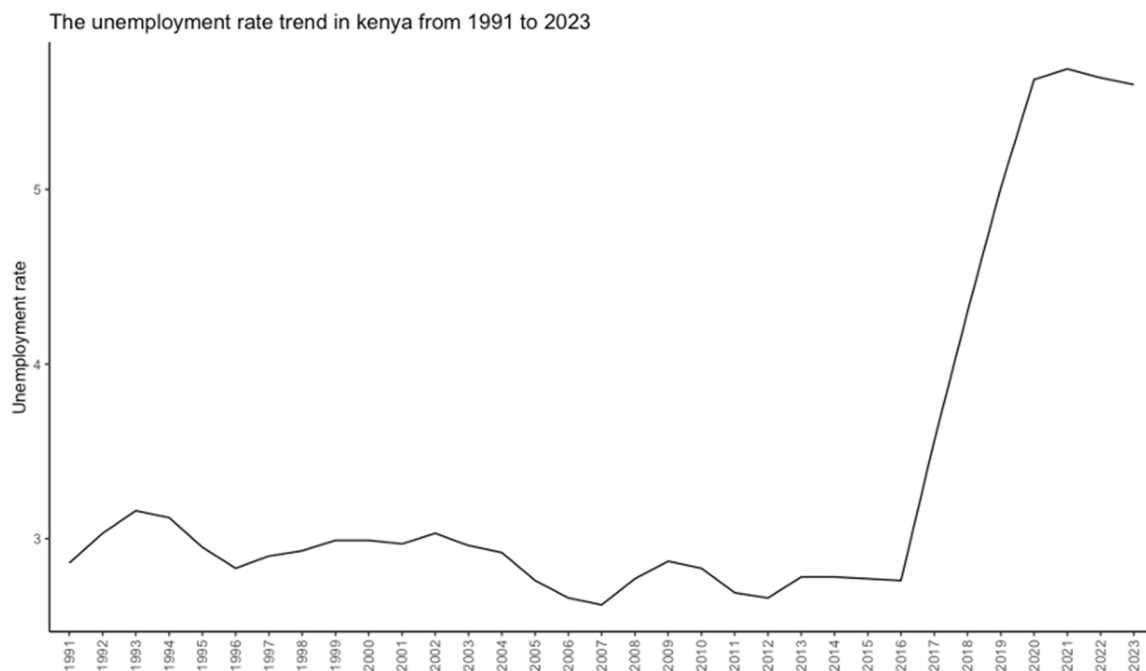


Figure 1. The unemployment rate trend in Kenya from 1991 to 2023.

The Figure 1 illustrates the trend in the unemployment rate in Kenya over a period spanning from 1991 to 2023.

In the early 1990s, the unemployment rate in Kenya fluctuated moderately, starting at approximately 3% in 1991 and rising slightly above 3% in the subsequent years. The early part of the decade shows minor fluctuations, with a slight increase around 1993, where the unemployment rate peaked close to 3.5%. However, from 1995 to 2000, the unemployment rate remained relatively stable, oscillating around 3%. The period from the mid-2000s to the late 2010s is marked by a stable unemployment rate with minor fluctuations. During this time, the unemployment rate hovered

around 3%, with slight increases and decreases observed between the years. Notably, from 2007 to 2010, there was a small peak where the unemployment rate rose slightly above 3% before declining again. As shown, there was a major shift in the type of trends surfacing from 2017: The previously stable unemployment rate started to climb swiftly. This is a significant rise since 2017 when unemployment was stable at (roughly) the 3% mark up to over approx. The sharp increase keeps going up to the highest point of this period by 2022, then it falls just slightly in 2023. The flatness of the unemployment rate from the early 1990s to late 2010 signalled a period of economic stability with stable low-level

disturbances. The short-term rise after 2017 may have been influenced by economic policies, global economic factors or other socio-economic issues such as government stability and lead to profound changes in the economy. The sharp rise in unemployment between 2017 and his previous year could also trace back to global economic conditions such as COVID-19, which caused mass disruptions of the world economy, sending worldwide jobless rates on to more extreme breaks. The very small decrease in 2023 may indicate a recovery has already started, or appropriate unemployment alleviation action being implemented.

3.2. Correlation Analysis

It is a statistical method used to assess the strength and direct of association between two numerical variables. This allows to comprehend how a variable can predict or be related with others which may have importance in different realms. In this analysis, we will concentrate on The Pearson Correlation Coefficient which is a measure of the strength and direction of association that exists between two continuous variables. A value of 1 represents a perfect positive linear relationship, -1 stands for a total negative linear correlation and the zero marks no linearity at all.

Table 3. Correlation analysis.

| Indicator | Unemployment Rate Percent |
|---------------------------------|---------------------------|
| Real GDP growth | 0.0783 |
| Gross public debt % of GDP | 0.8417 |
| Population | 0.6087 |
| Government revenue % of GDP | 0.1688 |
| Government expenditure % of GDP | 0.5040 |

The correlation between the unemployment rate and real GDP growth is 0.0783. This positive but weak correlation suggests a minimal relationship between unemployment rates and GDP growth in the countries analysed. It indicates that changes in GDP growth have little to no direct effect on unemployment rates. The correlation between the unemployment rate and gross public debt as a percentage of GDP is 0.8417. This strong positive correlation indicates that higher levels of public debt are associated with higher unemployment rates. This relationship suggests that countries with substantial public debt might face economic challenges that contribute to higher unemployment. The correlation between the unemployment rate and population is 0.6087. This moderate positive correlation indicates that larger populations are associated with higher unemployment rates. This relationship may reflect the challenges larger populations face in providing sufficient employment opportunities for all citizens. The correlation between the unemployment rate and government revenue as a percentage of GDP is 0.1688. This weak positive correlation indicates a slight association between higher government revenues and higher unemployment rates. The correlation with the unemployment rate and Government expenditure as a percentage of GDP

is 0.504. Well, it means that government spending tends to move in the same direction as unemployment, but not by too much. This correlation might have to do with fiscal policies, i.e., the desire of more government spending as a response to higher unemployment. This moderate correlation points to the fact that ideals such as government spending are partially externally-controllable over these rates, while other factors also play a role.

3.3. Multiple Linear Regression Modelling

Multiple Linear Regression (MLR) is a statistical technique used to model the relationship between one dependent variable and two or more independent variables. It is an extension of simple linear regression in order to model and analyse more complex behaviour between the dependent variable (let us say, y) with multiple number of independent variables or predictors.

Multiple linear regression formula:

$$\hat{Y}_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \dots + \beta_k X_k + \epsilon_i \quad (53)$$

Values:

1. \hat{Y} refers to the outcome value.
2. β_0 is the intercept value, the predicted value of \hat{Y} when main predictor variables (X_i) are constant.
3. $\beta_1, \beta_2, \dots, \beta_k$ are the coefficients of the independent variables X_1, X_2, \dots, X_k . Each coefficient represents the change in the dependent variable for a one-unit change in the corresponding independent variable, holding all other variables constant.
4. X_1, X_2, \dots, X_k are the predictor variables, the predictors or factors that might influence Y .
5. ϵ refers to the residual term, representing a portion of Y not explained by the model; it captures all other factors that influence Y but are not included in the model.

Table 4. Multiple linear regression result.

| Dependent variable: | |
|-----------------------------------|------------------------|
| Unemployment rate (%) ~. | |
| Real GDP growth | −0.008 (0.034) |
| Gross public debt (% of GDP) | 0.065*** (0.008) |
| Population | 0.104*** (0.018) |
| Government revenue (% of GDP) | −0.098* (0.051) |
| Government expenditure (% of GDP) | −0.140*** (0.033) |
| Constant | 0.892 (0.649) |
| Observations | 32 |
| R-Squared | 0.881 |
| Adjusted R-Squared | 0.858 |
| Residual Std. Error | 0.345 (df = 26) |
| F Statistic | 38.349*** (df = 5; 26) |

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The purpose of the multiple linear regression analysis is to investigate which economic indicators are related with unemployment rate percentage. The dependent variable was

the unemployment rate percentage and independent variables were real GDP growth, gross public debt to GDP ratio, population; government revenue to as a percent of GDP, government expenditure as a % of GDP.

There were no significant bivariate correlations. The results of the regression showed that real GDP growth was not significantly predicted to unemployment rate, $B = -0.008$, $SE = 0.034$, $p > 0.1$. It indicates that variations in GDP growth cannot be statistically shown to cause the unemployment rate. Instead, a more detailed exam shows that gross public debt (% GDP) was indeed the strongest positive predictor of unemployment rate with $B=0.065$, $p < 0.01$. Thus, a rise in public debt is linked to increased rates of unemployment. Within the positive predictors, population also was a major stimulant for success ($B = 0.104$, $SE = 0.018$; $p < 0.01$, indicating that population growth is connected to increases in the unemployment rate. An increases in population may be associated with higher levels of competition for jobs, which leads to more unemployment.

Government revenue (% of GDP): This variable had a significant negative effect on the unemployment rate, $B = -0.098$, $SE = 0.051$, $p < 0.10$. This is an economically significant relationship at the 10% level and shows that higher government revenue leads to lower unemployment. This suggests that government funded revenue generation through the likes of taxes (eg. negative coefficient) might create a pool from which to execute on public services and jobs activities, hopefully resulting in less unemployment as above. Likewise, government expenditure in % of GDP was a strong negative predictor $B = -0.140$ ($SE = 0.033$), $p < 0.01$. This implies that higher government expenditures are strongly negatively related to the unemployment rate. The robustness of the results remained intact. Among the many things it shows is how important government spending can be to employment,

pointing out that more public investment in infrastructure and social programs and other forms of increased expenditure can help put people back to work. Constant term: $B = 0.892$, $SE = 0.649$, $p > 0.1$ indicating that if all predictors are zero, then the baseline unemployment rate is not significantly different from zero. This means simply taking the average unemployment rate cannot accurately predict what will happen in other years without also looking at those independent variables.

Result for percentage of variation in unemployment rates that were explained by the model $R\text{-Squared} = 0.881$, Adjusted $R\text{-Squared} = 0.858$, indicating a strong fit. This implies that around 88.1% of the variance in unemployment can be foreseen from the model. A high adjusted $R\text{-Squared}$ value (adjusting for the number of predictors used in the model) indicates that this is a strong and robust model. The model was overall statistically significant, $F(5, 26) = 38.349$, $p < 0.01$, confirming that predictions systematically account for within state variance in the level of unemployment. Since the F -statistic is relatively large, we can conclude that our model fits significantly better to the data than a null (i.e., no predictors) model.

3.4. Normality Test

Table 5. Results of the Shapiro-Wilk Normality Test.

| Statistic | P-Value | Method |
|-----------|-----------|-----------------------------|
| 0.9690895 | 0.4746534 | Shapiro-Wilk normality test |

A Shapiro-Wilk normality test was conducted to assess the normality of the residuals. The results indicated that the residuals were normally distributed, $W(33)=0.97$, $p=0.47$.

3.5. Machine Learning Techniques

Table 6. Comparison of Modern Regression Models.

| Model | MAE | RMSE | R-Squared | AIC | BIC |
|---------------------------------|--------|--------|-----------|-----------|----------|
| Support Vector Regression (SVR) | 0.0823 | 0.0878 | 0.9915 | -100.2072 | -69.4268 |
| Neural Network Regression | 0.0881 | 0.0988 | 0.9887 | 291.8910 | 614.3529 |
| Bayesian Regression | 0.2579 | 0.3109 | 0.8806 | 28.0408 | 36.8352 |

SVR has registered its exceptional performance with Mean Absolute Error (MAE) of 0.0823 and Root Mean Squared Error (RMSE) of 0.0878, empathising pinpoint accuracy almost to the errorless predictions in several folds. Its $R\text{-squared}$ of 0.9915 shows that this model gave an nearly perfect fit, indicating close to the %99.15 percent variance in dependent variable can be explained by Independent Variables using this linear regression Model. It is striking to note that the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), were -100.2072 and -69.4268, respectively; indicating this excellent fit model

better than normal regression outcomes Neural Network Regression yielded an MAE: 0.0881, RMSE: 0.0988 with these numbers being significantly high; though not as good as SVR model. This is commendable as $R\text{-Squared}=0.9887$ indicates the model explained approximately 98.87% variability in dependent variable. However, having both an elevated AIC of 291.8910 and a very large BIC value of 614.3529 also signalled the poor statistical fitting with respect to SVR. The MAE value is obtained as 0.2579 and RMSE value for Prediction of the combined dataset was recorded to be - 0.3109. So on total, a decent performance can seen from

Bayesian Regression in predictions with R-Squared 0.8806. However, the model showed lower AIC and BIC values of 28.0408 and 36.8352, respectively.

4. Conclusion

According to the regression analysis of key models such as: SVR, Neural Network Regression, and Bayesian Regression, unique result trends came out. SVR performed well, with MAE of 0.0823 and a RMSE of 0.0878, showing excellent accurate predictions with less error. The SVR's R-Squared value of 0.9915 showed that it accounted for approximately 99.15% of the variance in the dependent variable indicates an almost perfect fit. The model is confirmed to be well fitted as indicated by the markedly low AIC value of -100.2072 and BIC value of -69.4268. On the contrary, Neural Network Regression had slightly lower accuracy with an MAE of 0.0881 and an RMSE of 0.0988 but still managed to score R-Squared = 0.9887. It has higher AIC and BIC values which are indicative of a worse fit because it has more parameters. Bayesian Regression had moderate accuracy, with an MAE of 0.2579 and an RMSE of 0.3109 while it scored R-Squared = 0.8806. Plus lower AIC and BIC values indicated a good fit that offered a balanced option between simplicity in model and accuracy.

Therefore SVR is recommended for high precision demanding situations due to its superior accurate fitting making it suitable for critical predictive tasks; however, Neural Network Regression may be preferable as it offers many advantages in handling complex relationships especially in large datasets albeit at some possible penalties on statistical efficiency found through reduced AIC or BIC scores. The Bayesian regression which was observed as giving moderate performance was considered as being suitable for less complex or non-critical prediction needs where computational efficiency is put above everything else such as model simplicity according to its smaller AIC and BIC values.

This research primarily concentrated on key factors determining the unemployment rate in Kenya, including Real GDP growth, Population, revenue by Government as a % of GDP, expenditure by Government as a % of GDP, and Gross public loan as a % of GDP. Additionally, the study utilised data spanning from 1991 to 2023. Therefore, for the purpose of future study more data need to be used including the latest one. Since the research focussed on the modelling, data prior to 1991 say from 1960 to current year can be used to train various models. In terms of determinants of the

unemployment rate in Kenya, future research could expand beyond the key determinants used in this study-such as Real GDP growth, Population, revenue by Government as a % of GDP, expenditure by Government as a % of GDP, and Gross public loan as a % of GDP-to include additional factors like: inflation rate, interest rates, workforce participation rate, educational attainment, demographic factors, government policies, consumer sentiment, global economic conditions and job vacancies

Abbreviations

| | |
|------|--------------------------------|
| AIC | Akaike Information Criterion |
| BIC | Bayesian Information Criterion |
| CPI | Consumer Price Index |
| GDP | Gross Domestic Product |
| MAE | Mean Absolute Error |
| MLR | Multiple Linear Regression |
| QTM | Quantity Theory of Money |
| RMSE | Root Mean Squared Error |
| SD | Standard Deviation |
| SVR | Support Vector Regression |

ORCID

0009-0001-4519-342X (Reuben Opondo)

0000-0003-3290-0444 (Davis Bundi)

0000-0002-6283-4567 (Patrick Weke)

Author Contributions

Reuben Opondo: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing

Davis Bundi: Formal Analysis, Methodology, Supervision, Visualization

Patrick Weke: Conceptualization, Data curation, Methodology, Supervision, Validation, Writing - original draft, Writing - review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

Appendix I: Economic Data from 1991 to 2023

Table 7. Kenya's Economic Data from 1991 to 2023.

| Year | ur | growth | gpb | pop | grp | gep |
|------|------|--------|-------|-------|-------|-------|
| 1991 | 2.86 | 1.30 | 43.04 | 22.70 | 9.88 | 16.44 |
| 1992 | 3.03 | -1.10 | 41.16 | 23.40 | 9.33 | 17.60 |
| 1993 | 3.16 | -0.10 | 61.64 | 24.10 | 10.35 | 18.92 |
| 1994 | 3.12 | 2.50 | 57.01 | 24.80 | 14.65 | 18.80 |
| 1995 | 2.95 | 4.30 | 52.08 | 25.50 | 17.10 | 17.33 |
| 1996 | 2.83 | 4.50 | 40.53 | 26.30 | 14.73 | 15.22 |
| 1997 | 2.90 | 0.40 | 35.98 | 27.00 | 14.64 | 15.55 |
| 1998 | 2.93 | 3.00 | 35.89 | 27.80 | 15.31 | 15.34 |
| 1999 | 2.99 | 2.20 | 38.41 | 28.70 | 14.52 | 13.68 |
| 2000 | 2.99 | 0.30 | 43.12 | 29.50 | 14.47 | 14.09 |
| 2001 | 2.97 | 4.00 | 41.28 | 30.30 | 14.41 | 14.94 |
| 2002 | 3.03 | 0.50 | 42.01 | 31.10 | 14.36 | 15.65 |
| 2003 | 2.96 | 2.90 | 43.79 | 32.00 | 15.29 | 16.01 |
| 2004 | 2.92 | 4.60 | 40.75 | 32.90 | 15.95 | 15.45 |
| 2005 | 2.76 | 5.70 | 37.41 | 33.80 | 16.06 | 16.25 |
| 2006 | 2.66 | 5.90 | 37.07 | 34.70 | 16.88 | 17.31 |
| 2007 | 2.62 | 6.90 | 34.19 | 35.70 | 17.12 | 18.08 |
| 2008 | 2.77 | 0.20 | 34.25 | 36.70 | 16.96 | 18.92 |
| 2009 | 2.87 | 3.30 | 35.96 | 37.70 | 17.15 | 20.27 |
| 2010 | 2.83 | 8.10 | 36.69 | 38.60 | 17.87 | 21.54 |
| 2011 | 2.69 | 5.10 | 35.69 | 39.50 | 16.48 | 20.12 |
| 2012 | 2.66 | 4.60 | 37.62 | 40.40 | 16.81 | 22.10 |
| 2013 | 2.78 | 3.80 | 39.76 | 41.40 | 17.98 | 23.23 |
| 2014 | 2.78 | 5.00 | 41.28 | 42.40 | 17.67 | 23.42 |
| 2015 | 2.77 | 5.00 | 45.83 | 43.30 | 17.14 | 23.83 |
| 2016 | 2.76 | 4.20 | 50.40 | 44.30 | 17.89 | 25.34 |
| 2017 | 3.56 | 3.80 | 53.87 | 45.40 | 17.80 | 25.17 |
| 2018 | 4.30 | 5.70 | 56.45 | 46.40 | 17.54 | 24.45 |
| 2019 | 5.01 | 5.10 | 59.08 | 47.60 | 17.00 | 24.40 |
| 2020 | 5.63 | -0.30 | 67.83 | 48.80 | 16.67 | 24.80 |
| 2021 | 5.69 | 7.60 | 66.97 | 49.75 | 16.82 | 24.02 |
| 2022 | 5.64 | 4.80 | 67.94 | 50.64 | 17.18 | 23.00 |
| 2023 | 5.60 | 5.50 | NA | 51.54 | NA | NA |

Appendix II: Data Dictionary

Table 8. Data dictionary.

| Code | Indicator | Unit |
|--------|--------------------------------------|--------------------|
| ur | Unemployment rate | Annual % change |
| growth | Real GDP growth | Annual % change |
| gpb | Gross public debt as a % of GDP | % of GDP |
| pop | Population | Millions of people |
| grp | Government revenue as a % of GDP | % GDP |
| gep | Government expenditure as a % of GDP | % GDP |

References

- [1] Barro, R. J., & Gordon, D. B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics*, 12(1), 101-121. [http://dx.doi.org/10.1016/0304-3932\(83\)90051-X](http://dx.doi.org/10.1016/0304-3932(83)90051-X)
- [2] Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999, January 1). Monetary policy shocks: What have we learned and to what end? *ScienceDirect; Elsevier*. [http://dx.doi.org/10.1016/S1574-0048\(99\)01005-8](http://dx.doi.org/10.1016/S1574-0048(99)01005-8)
- [3] Davis, S. J. (2012). Variations in the wake of labor reallocation. *Carnegie-Rochester Conference Series on Public Policy*, 27, 335-402. North-Holland.
- [4] Edward, S. (2007). How useful is Okun's law? *Economic Review*, 73-103.
- [5] Friedman, M. (2010). Quantity theory of money. *Monetary Economics*, 299-338. https://doi.org/10.1057/9780230280854_35
- [6] Hu, X. (2017). Support vector machine and its application to regression and classification.
- [7] Karatekin, K., Gocer, I., & Yazar, E. (2015). The relationship between youth unemployment and economic growth in Central and Eastern European countries.
- [8] Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, 58(4), 757. <https://doi.org/10.2307/2938349>
- [9] Mishkin, F. (2004). Why the Federal Reserve should adopt inflation targeting. *International Finance*, 7(1), 117-127. <http://dx.doi.org/10.1111/j.1367-0271.2004.00132.x>
- [10] Moses, V. (2019). Does Okun's law on cyclical unemployment apply in Kenya? *International Journal of Current Research*, 1765-1770. <http://dx.doi.org/10.2139/ssrn.2640519>
- [11] Pollin, R., Mwangi wa Githinji, & Heintz, J. (2008). An employment-targeted economic program for Kenya. Edward Elgar Publishing.
- [12] Ryan, T. C. I. (2002). Policy timeline and time series data for Kenya.
- [13] Seyfried, W. (2004). Examining the relationship between employment and economic growth in the ten largest states. *Southwestern Economic Review*, 1-12.
- [14] Sine, W. D. (2010). Institutions and entrepreneurship. *RESEARCH in the SOCIOLOGY of WORK*, 21, 1-26. [http://dx.doi.org/10.1108/S0277-2833\(2010\)0000021017](http://dx.doi.org/10.1108/S0277-2833(2010)0000021017)
- [15] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222. <https://doi.org/10.1023/b:stco.0000035301.49549.88>