# SPATIAL REGRESSION OF THE GROSS COUNTY PRODUCT OF KENYA ON INDUCED LATENT VARIABLES

**SAMSON MNAZI TSUMA**

**A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for the Master of Science Degree in Statistics of Egerton University**

**EGERTON UNIVERSITY**

**OCTOBER 2024**

## DECLARATION AND RECOMMENDATION

### Declaration

This thesis is my original work and has not been presented in this university or any other for the award of a degree.

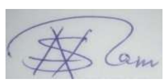
Signature: ……………………………… Date: …………01/10/2024…………

Samson Mnazi Tsuma

…SM123/14726/18…………..…………

### Recommendation

This thesis has been submitted with our approval as university supervisors.

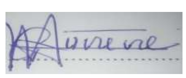


Signature: ……………………………… Date: ………01/10/2024………………

Prof. Ali Salim Islam

Department of Mathematics

Egerton University



Signature: ………………………………… Date: …………01/10/2024……………

Dr. George Matiri

Department of Mathematics

Egerton University

## COPYRIGHT

# DEDICATION

I dedicate this work to all the scientists who came before me, opening the path for my research in this field. May this add up to the mutual pursuit of knowledge.

**ACKNOWLEDGEMENTS**

I would like to appreciate my supervisors Prof. Ali Salim Islam and Dr. George Matiri for their support and guidance throughout the entire duration of this research. Their expertise and passion for the subject matter greatly influenced the quality and direction of this thesis. I am also grateful to the member of my thesis committee Dr. Justin Obwoge for his insightful input and fruitful discussions during the various stages of this research. His expertise and valuable suggestions significantly contributed to the refinement of this work.

Lastly, i extend my sincere appreciation to the staff of Egerton University for their assistance and resources provided which facilitated the smooth execution of this study. Their willingness to support and accommodate my research needs played a crucial role in the successful completion of this thesis.

## ABSTRACT

Since the re-organization of the local boundaries in Kenya from provinces to counties in 2013, there had been few studies carried out to measure regional economic progress. The role of geographical analysis in economic development had been of less concern in Kenya. Investigating the spatial dependence of the Gross County Product GCP of Kenya on latent variables was important as it solved the error of model misspecification and in the relationship could have proved the spill-over effect of the Kenyan economy at the county levels. This cross-sectional study identified the common factors among the economic indicators through factor analysis and the economic development of the country by use of thematic maps. The Kenya National Bureau of Statistics (KNBS) published and publicized the 2019 geocoded dataset as a Kenyan economic survey report for that year which consisted of 47 Kenyan counties and 18 county economic indicators which were used in this study. The Local Indicator of Spatial Association (LISA) (Moran I test) for spatial clustering and the Maximum Likelihood Estimation (MLE) method was employed to obtain the parameter estimates of the spatial relationship which were important for policy making among various economic stakeholders in the country. The Lagrange Multiplier (LM) Test together with the spatial Hausman test suggested an error model fit, lags being significant, or endogeneity. Meanwhile, the likelihood ratio test considered a restricted spatial model more suitable than the nested model. The ArcGIS (version 10.7.1) and R (version 4.3.0) software were used for spatial analysis. The Spatial error model (SEM) results gave out a suitable equation that revealed the relationship between the GCP and the county indicators. This produced a benchmark for explaining geocoded datasets for monitoring the economic pattern of Kenya and correct how the 6 economic blocs were classified geographically while highlighting the drawbacks in the achievement of Kenya's development programs like Vision 2030 and Sustainable Development Goals.

TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS AND ACRONYMS

**FA** Factor Analysis

**GCP** Gross County Product

**GDP** Gross Domestic Product

**GNSM** General Nesting Spatial Model

**GRDP** Gross Regional Domestic Product

**KADP** Kenya Accountable Devolution Program

**KNBS** Kenya National Bureau of Statistics

**LM/RS** Lagrange Multiplier/Rao's Score test

**MLE** Maximum Likelihood Estimation

**MRM** Minimal Residual Method

**NUTS2** Nomenclature of Territorial Units for Statistics Level 2

**NUTS3** Nomenclature of Territorial Units for Statistics Level 3

**OLS** Ordinary Least Squares Model

**PCA** Principal Component Analysis

**SACM** Spatial Autoregressive Confused Model

**SAR** Spatial Autoregressive Model

**SARAR** Spatial Autoregressive Model with Spatial Disturbances

**SDM** Spatial Durbin Model

**SDEM** Spatial Durbin Error Model

**SEM** Spatial Error Model

**SLX** Spatially Lagged X Model

# CHAPTER ONE
# INTRODUCTION

## 1.1 Background information

Regional economic reports for a country are commonly used for discussing disaggregated economic performance values that add up or relate to the total domestic product. For example, the annual economic reports in South Africa were given in the form of a regional breakdown instead of a single value for the country's GDP. For the first time in Kenya, this was manifested in 2019 when the Kenya National Bureau of Statistics (KNBS) with the help of the World Bank created the GCP through the Kenya Accountable Devolution Program (KADP) and gave an expounded report in the KNBS (2019) for monitoring regional economic status. The GCP is a measure and assessor (proxy) of economic progress just like the Gross Domestic Product (GDP) but at county levels. The assembling of the values was according to the international guidelines for the estimation of regional gross domestic product. The process involved the identification and verification of suitable indicators that accurately revealed levels of economic activities for the various sectors at the county level. These indicators were then used to divide the overall GDP into GCP for the various counties. The sum of the GCP is equal to the national-level GDP. KNBS's (2019) reason for these indicators was to give a picture of the economic structure and relative size of the economy for each county.

The functions of the GCP included; Estimation of revenue potential for each county, informing economic progress at the county level, indicating the potential for private sector investment, and informing on county economic development plans. There were 18 indicators used for the GCP calculation which are in details as shown in appendix A and they included education (Educ), Agriculture, forestry and fishing (AgrForFish), Mining and quarrying (MinQua), Manufacturing (Manuf), Water supply; waste collection (WatSuWCol), Construction (Constr), Wholesale and retail trade; repair of motor vehicles (WholRRMV), Transport and storage (TranSto), Accommodation and food service activities (AccoFodSA), Information and communication (Infocom), Financial and insurance activities (FinInsuA), Real estate activities (RelEstA), Professional, technical, and support services (ProTecSupS), Public Administration and Defence (PubAdminD), Electricity supply (ElectSup), Human health and social work activities (HumHelSocWA), Other service activities (Other services), and Financial intermediation services indirectly measured (FinIServM1).

There had been a slow progress on spatial modeling in Kenya at regional levels that defined a spatial economic process since Okwi *et al*. (2007) checked on poverty at provincial levels. The use of simple linear models for economic assessment had produced unsatisfactory results that did not consider the geographical and boundary effects. According to McMillen (2012), spatial modeling provided convenient model specification tests for a spatial relationship that was not adequately reflected by the linear model. In this study, the spatial effect of the factors that affect the GCP of Kenya was examined through spatial regression modeling.

Ramirez and Loboguerrero (2002) disclosed that spatial effects were chief in spelling out economic growth patterns. These effects included spillover and externalities across the economies as observed in Balash *et al.* (2020). Spatial interrelation in the analysis of economic growth if side-lined could have resulted in an inaccurate model. Because the GCP was geographically oriented, the disaggregation concept was used to find out if there existed a spatial relationship and marginal contributions from affecting values which were the factor score results of the indicators mentioned in the previous paragraph. Each of the 47 counties of Kenya produced values in millions of Kenya shillings at both constant and current prices of goods and services however this production could not be said to be exclusively from a single county. Consequently, there was the cross county dependence in terms of each indicator of production. This dependence across the counties necessitated the use of spatial models to model the GCP.

Spatial autocorrelation was used to capture the spatial dependence or clustering as discussed by Moran (1950). For example, the fishing activity in Lake Victoria affected counties like Busia, Kisumu, Homabay, and Migori. Therefore, the GCP for each of the counties could not be explained separately because of the presence of a shared water body. From Burkey (2018), the popular spatial regression models that had been used in literature included the Spatial Error model (SEM), Spatial Lag Y (Autoregressive) Model (SAR), Spatially Lagged X (SLX), Spatial Durbin Error model (SDEM), Spatial Durbin Model (SDM) and the General Nesting Spatial Model (GNS)(Manski all-inclusive model).

Spatial modeling of the GCP was an important concept that advanced and clarified the impact of government, public, and private sectors on economic activities in the country of Kenya. Grouping of the GCP indicators through factor analysis enabled the stakeholders of the included activities to focus on development under clustered objectives while obtaining clarification on the spatially classified Kenyan economic blocs.

### 1.2 Statement of the problem

The research on geographical modeling for county economic planning and policy making in Kenya had not been thoroughly explored. Spatial models had evidently not been popularized to observe the relationship of the geographical effect with econometric variables for economic planning among the counties of Kenya. The Gross county product was a new phenomenon that depended on the economic indicators and how much each contributes to the relationship. Economic policy making from a geographical viewpoint became complex and misinterpreted when using non-spatial models (model misspecification). During the compilation of the GCP, the indicators were directly summed up without considering whether the indicators were highly correlated or not. The effect of geographical points representing the county from which the indicators came from was not included in the model formed. Furthermore, due to the large set of indicators, there was vulnerability of making dimensional errors during interpretation of multiple regression results.

The utility of regional econometric research that employed data reduction methods and the use of latent variables had not been greatly performed in Kenya. County economic progress was majorly considered exclusive by majority of people. The GCP 2019 results neglected the marginal effects that could have shown the impacts of counties neighboring each other. While one of the functions of the GCP was to inform on the county development plans, it was important to understand that efforts on development were majorly observed from a geographical model of spatial dependence which was not clearly known among many scholars. Thus, there was a gap in Kenya on spatial modeling with respect to counties. The February 2019 GCP results were only given as a general report and though a lot of money was used in their generation, there had not been any further advantages seen from them. They could have been a correct necessity if the spatial analysis that was to bring forth insight and county development policies could be implemented.

### 1.3 Objectives

#### 1.3.1 General objective

To investigate the effects of the geographical location on the relationship between the GCP of Kenya and its indicators giving rising to an econometric framework.

#### 1.3.2 Specific objectives

i. To identify the most important key GCP indicators using factor analysis so that multi-dimensional errors are minimized in the spatial regression modelling.

ii. To determine the best spatial regression model that relates GCP to its most important key indicators.

iii. To determine whether the marginal effect of the induced variables on the GCP was present to show the impact of dependence between neighboring counties.

iv. To draw the thematic maps of the identified most important key indicators of GCP of Kenya for presentation, cadastral purposes and future references.

### 1.4 Justification

The KNBS (2019) GCP report was not a complete reflection of the underlying concept of the regional economic relationship between the GCP and the economic indicators. The mathematical model that was used to unveil these had no parameters that expressed the geographical locations recorded in the dataset. Moreover, the direct summation of the indicators was a miscalculation due to the avoidance of dimensional and collinearity concepts. Therefore, a spatial regression analysis was conducted to deal with the imprecise manner of defining the economic evolution. The Maximum likelihood estimates from a spatial relationship simplified the complex system capturing the marginal (direct or indirect) or a spill-over effect. It was of great fault to have an exclusive analysis on the economic production for each Kenyan county without considering the impact and contribution of the counties neighboring each other. The specification on the spatial models by the autocorrelation effect revealed the stochastic shock unidentified by simple linear models. A conclusive nature which was a remedy for the errors of excessive dimensions from the Gross county product indicators was induced through factor analysis and the resulting latent variables. This brought about a change in the country's economic policies at county levels as a disaggregate procedure. Factor analysis evaluated the counties economic development by classifying and simplifying parameters after extracting a number of common factors from which ranks were calculated with respect to scores. The dimension reduction enabled the 18 indicators of measuring economic transition to be grouped and provided descriptive summary at county levels. From this, we obtained the induced latent variables that gave us the marginal effects on the GCP. The geographical classification of the counties from the economic dataset revealed the need to have new spatial economic blocs in Kenya and focus on their development.

Robustness checks in spatial regression models provided a strong foundation on important statistical outcomes and furthermore accommodated the class of models in which the main reason was to estimate the causal dependence of neighboring values of the dependent variable on itself. The existence of a stochastic shock in space made this cross-sectional study crucial and necessary as it consisted of a mathematical and statistical derivation procedure of spatial estimates that defined in details about the economical spatial relationship.

The KNBS (2019) report had not considered the maps for all indicators and thus generalized on the economic condition of the counties. The thematic maps for the indicators in this study were of great benefit to stakeholders as they touched on spatial information not only in the form of values but also as a graphical presentation. For example, quality education and clean water for sanitation were a part of the Sustainable development goals (SDG's) which are to be achieved by 2030 and also they were variables that affected the Gross county product that was to be mapped county-wise. The implication of spatial modeling, thematic mapping and factor analysis of the Gross county product contributed greatly to the Kenyan economy both practically and theoretically. The economic problems at county levels were easily be monitored and zoomed into for effective problem solving and policy resolutions by the government that was trying to balance its performance evenly and in the society as a whole. Therefore, the results of the study gave future directions on modelling regional economic associations and highlighted the progress on the set development blueprint.

**1.5 Assumptions and Limitations of the study**

i. The Gross county product was an estimate added at county levels to quantify the relative economic size in Kenya.

ii. The cross-sectional population data was well defined and therefore all potential bias were eliminated.

iii. Since there were large matrices and vectors used in the regional analysis that cannot even be displayed in the research, it was hard to discuss an array of values unless further computations which are tiresome.

# CHAPTER TWO
# LITERATURE REVIEW

## 2.1 Overview of spatial models

Spatial econometrics was a topic that worked on geographical spatial interaction (spatial autocorrelation) and spatial locational structure (spatial heterogeneity) in regression models for cross-sectional and panel data Paelinck (1978). Spatial models had been applied by many researchers especially geologists and epidemiologists. For example, research on spatial effects had been a progressive aspect throughout the world as observed by Moreno and Trehan (1997) who showed that among 93 nations a country's growth rate was positively influenced by the growth rate of neighboring countries through the spill-over effect at their borders. The new concept of the Kenyan GCP in 2019 brought an urge to utilize spatial econometric model techniques through this study as an upgrade in regional spatial modeling.

Xiangyu *et al*. (2018) considered whether to use the Spatial Durbin model (SDM), spatial lag Y model (SLM) or Spatial error model (SEM) to discuss about the nitrogen oxides emissions amount at China's provincial level that were being influenced by log transformed variables like the total population and energy intensity. He identified a spill-over effect of nitrogen oxides emissions among the neighboring provinces through the Spatial Durbin Model when the polynomial concept was adopted in a nested cubic model. The SDM could not be simplified to the SLM or SEM as concluded by the likelihood ratio (LR) test. The significant spatial spillover effects of nitrogen oxides emissions suggested that policymakers, especially local governments, were to not only focus on the local emission level but also consider the influence of the neighboring provinces.

Karim *et al.* (2017) modelled the Gross Regional Domestic Product (GRDP) of Bruto in Central Java Province, Indonesia using Spatial Regression. In his study, factors that influenced the GRDP such as human capital gave significant influence in the linear and SLX model while in the SLX model only the weighted variable of labor had significant effect. The best model was the SLX with an R-squared value of 0.64. Thus, the conclusion was that the GRDP value in a region in Central Java was influenced by the value of the human capital of the region as well as the labor of the nearest region (local spatial model).

Szendi (2016) discussed about the spatial beta convergence analysis of the real GDP per capita across Germany and Hungary. In the analysis, she examined the beta convergence with special regards on spatial processes. The field of the analysis was the 40 NUTS2 regions and the 434 NUTS3 districts of Germany and the 20 NUTS3 territories (counties and the capital) in Hungary. The applied indicator of the analysis was the GDP per capita of the territories based on power parity standard in the period of 2000 to 2013. The results were such that in the case of the Hungarian counties the beta divergence was realized between 2000 and 2013, and the spatial effects were non-significant.

In the German NUTS2 regions there was a realizing beta convergence in the analyzed time period, but the spatial effects were also non-significant. The case of the German NUTS3 districts was a kind of special, because here beside the beta convergence of the territories, it was observable a significant spatial lag model, which was fitting better than the linear model. In this case the regression contained a spatially lagged dependent variable which had got positive, but only weak effects on the convergence.

Okwi *et al*. (2007) came up with a provincial spatial relationship of poverty in Kenya for both exogenous and endogenous variables that explained the welfare levels in different areas within provinces. The exogenous variables included geographic factors such as rainfall and the endogenous ones were majorly demographic factors. The discussion angled on the spatial lag Y and spatial error dependence at both provincial and national levels. Results of the regression models demonstrated the statistical significance of certain spatial variables. At provincial levels, the variables employed were heterogeneous and important for designing and evaluating provincial-specific poverty-reduction strategies. The analysis helped to explain the geographic determinants of poverty.

### 2.2 Application of factor analysis in econometric studies

Factor analysis had been used widely in Gross domestic product (GDP) analysis in several countries including Bangladesh and Pakistan.  For example, in Pakistan three major factors that influenced the GDP were found from the analysis according to Syed *et al*. (2013). The first factor explained the service related activities in industrial and business of the country while the second factor is purely dominated by agricultural and livestock sectors and the last factor is purely dominated by agricultural and livestock sectors. There were seventeen variables of the gross district products of 64 districts in this study for the year 2010-2011.

In Bangladesh, the analysis from Akhter *et al*. (2012) revealed that seventeen sectors had been classified into three factors that are contributing to Bangladesh's GDP. These three factors for principal component analysis were renamed as service factor, agriculture and infrastructure factor, and fishing together with mining factor. Since the availability of gross domestic product data was very scarce for older days, the data for the year 1999-2000 was used for the analysis.

Bai *et al*. (2015) evaluated the country economic rank by applying factor analysis using International monetary fund (IMF) dataset for 20 countries. The result showed the economic rank of countries (Kuwait, Germany, Iceland, Belgium, Denmark, Taiwan, Qatar, Ireland, Sweden, Luxemburg, Austria, Singapore, Norway, Netherland, Hong Kong, Brunei, US, Switzerland, Canada, and Australia). Also, the calculated rank and the rank provided by world ranking list was almost the same which confirmed that it was successful to apply factor analysis into countries economic evaluation. The 2015 paper described the basic principles of factor analysis, and used the method to perform a comprehensive analysis and evaluation of economic development of 20 countries on 21 economic parameters.

In this case, we had the context of the GCP which from comparison with the listed literature, was a new regional concept applied uniquely to Kenya. Since there were 18 economic indicators from the 2019 GCP, factor analysis was the key concept to extract the common variances and reduce them into fewer number of factors. The latent variables that were created were then used in the spatial regression to form the latent spatial regression models as suggested by Oud and Folmer (2008).

**2.3 Application of spatial econometric models**

Spatial models had their roots from the simple linear regression model. Spatial dependence in a collection of sample data implied that;

$$\underline{y} = f(\underline{X}, \underline{\beta}) + \underline{\varepsilon} \,. \qquad (2.1)$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots \ldots \ldots .. + \beta_n x_{in} + e_i \,.$$

where i = 1,…….,n and $\underline{y}$ was the dependent variable denoted as an n ×1 vector of observations on the outcome variable, $\underline{X}$ was an n × m matrix for n observations of m independent variables, $\beta$ was an m ×1 vector of regression coefficients and $\underline{\varepsilon}$ was an n ×1 vector of random errors as discussed by Fotheringham *et al.* (2000). The dependence was allowed to be among several observations, as the location index $i$ took on any value from$i = 1, \ldots, n$. The model in equation (2.1) was often criticized for not revealing the spatial and temporal effects. In this particular discourse, the concern will be to observe the residual pattern of the model in equation (2.1) if spatial effects existed.

The seven spatial models which capture the spatial effects included:

**2.3.1 Spatially lagged X (SLX)**

The SLX model was given as;

$$Y = X\beta + \theta(I_T \otimes W_N)X + (\tau_T \otimes I_N)\varepsilon + (I_T \otimes \tau_N)\eta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I_{NT}).$$

$$y = X\beta + WX\theta + \varepsilon. \tag{2.2}$$

This model was a local spatial model (revealed the effects of the neighbors on every point in space) that added the average values of the neighboring X's. The idea was that the effect of the independent variables was not only in the region with the points in space but also the neighboring regions. Any predictor variable could have had an effect on the same response factor in the neighboring region that could have spread to the points in space. The $k \times 1$ parameter vector $\theta$ measured the marginal impact of the explanatory variables from neighboring observations on the dependent variable y.

$W$ was an (n×n) non-stochastic spatial weights matrix, with zeros on the main diagonal and non-negative values for $W_{ij}$ , i ≠j. Multiplying X (n×k) matrix of exogenous explanatory variables by W produced spatial lags of the explanatory variables that reflected an average of neighboring observations X−values. W was the row-stochastic matrix. Thus, we expressed the utility, y as a function of observable characteristics Xβ, the spatial lags of the explanatory variables WX and the unobservable characteristics ε.

Huang (2014) suggested the existence of spatial lag dependence due to the presence of social and spatial interactions over three periods. His research focused on energy demand which had a spatial lag dependence on negative price elasticity, positive but declining income elasticity and the significant effects of industry/service value added, urbanization and technical innovations.

Kim *et al.* (2003) also engaged the spatial-lag model as a housing price model for the Seoul metropolitan area to measure the marginal value of improvements in sulfur dioxide ($SO_2$) and nitrogen dioxide ($NO_2$) concentrations through diagnostic testing. His Results showed that $SO_2$ pollution levels had a significant impact on housing prices while $NO_2$ pollution did not.

### 2.3.2 Spatial lag Y (Autoregressive) model (SAR)

The SAR model was given as;

$$Y = \rho(I_T \otimes W_N)Y + X\beta + (\tau_T \otimes I_N)\varepsilon + (I_T \otimes \tau_N)\eta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I_{NT}).$$

$$y = \rho W y + X\beta + \varepsilon \; . \tag{2.3}$$

This was a global spatial model (revealed the spatial effects of not only the neighbors but also the spill over to the other regions) in which an endogenous variable was specified to depend on spatial interactions $\rho Wy$ between cross-sectional units plus a disturbance term. The interactions were modeled as a weighted average of nearby cross-sectional units, and the endogenous variable comprising the interactions was usually referred to as a spatially lagged variable. The weights (W) were grouped in a matrix identifying neighborhood connections, which forms the distinctive core of spatial process models. The elements of W were specified before the model was fit. If the matrix had contained all 0s, there would have been no spatial effects and we would as well just fit the linear regression model.

The model was termed as spatial autoregressive lag model in the terminology of Anselin and Florax (1995). Whittle's spatial autoregressive lag model (SAR) was popularized and extended by Cox (1984), who distinguished models in which the disturbances followed a spatial autoregressive process in Karim *et al.* (2017). If autocorrelation was embedded in the dependent variable, the spatial autoregressive (SAR) model with a spatial lag could be used. SAR models extended the linear regression by allowing outcomes in one area to be affected by outcomes in nearby areas, covariates from nearby areas and errors from nearby areas.

We did not have the spatially lagged X's but the spatially lagged Y's in this case. If there were higher effects of the response variable in the neighboring areas then it may have been spilling over from the points in space and also beyond.

The SAR model had been applied widely in research like by Pozdílková and Marek (2018) who described the geo-informational phenomena of the housing prices in small municipalities within Pardubice region in Czech Republic from a spatially lagged model approach.

Kanaroglou *et al*. (2013) also estimated the sulfur dioxide air pollution concentrations in Canada using this model and obtained an improvement in the pseudo R squared, log likelihood and reduced mean squared error as compared to the base model.

### 2.3.3 Spatial error model (SEM)

When there was a residual pattern in the error component of the traditional regression model, the spatial error model (SEM) could have been used. It was also a global spatial model given by;

$$Y = X\beta + (\tau_T \otimes I_N)\mu + (I_T \otimes \tau_N)\eta + \mu.$$

$$\mu = \lambda(I_T \otimes W_N)\mu + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_{NT}).$$

$y = X\beta + \mu.$ and $\mu = \lambda W\mu + \varepsilon.$

making $\mu$ the subject by substitution we had;

$$y = X\beta + (I_n - \lambda W)^{-1}\varepsilon \ , \ \varepsilon \sim N(0, \sigma^2). \qquad (2.4)$$

W was the spatial weight matrix representing the structure of the spatial relationships between observations. $\lambda$ was the spatial dependence parameter and $\mu$ was a vector of auto correlated disturbances, and all other terms were the elements commonly found in ordinary linear regression analysis. The ideology was that the interaction effect among the error terms $\lambda W\mu$ that was spread over the regions was from the residuals $\mu$ i.e. unexplained value which was a function of the neighbor's residual values plus the stochastic error term.

Bai En Chong *et al.* (2012) investigated on the spatial structure of the provincial economic growth and the spatial spillover in China from 1998 to 2008 using the SEM to account for the spatial autocorrelation. This was the same model that Okwi *et al*. (2007) based his findings on poverty at provincial levels in Kenya.

### 2.3.4 Spatial Durbin error model (SDEM)

This was a local spatial model with effects only on the direct neighbors. It was given by;

$$Y = X\beta + \theta(I_T \otimes W_N)X + (\tau_T \otimes I_N)\mu + (I_T \otimes \tau_N)\eta + \mu.$$

$$\mu = \lambda(I_T \otimes W_N)\mu + \varepsilon \ \varepsilon \sim N(0, \sigma^2 I_{NT}).$$

$$y = X\beta + WX\theta + \mu. \qquad (2.5)$$

where $\mu = \lambda W\mu + \varepsilon.$

Afifah *et al.* (2021) using the SDEM model, showed that there was a spatial relationship where the amount of tax revenue in each region was different and was influenced by other areas in East Java. Septiawan *et al.* (2018) used the SDEM as he observed that the model overcame the spatial effect of errors and the effects of spatial dependency on the independent variable for the Human Development Index in Central Java Province, Indonesia. Jiang *et al.* (2017) analyzed the determinants of energy efficiency by means of an SDEM for 29 Chinese provinces over the period 2003–2011 by considering both factors in their own province and in first-order neighboring provinces.

### 2.3.5 Spatial Durbin model (SDM)

The spatial Durbin was the model containing both endogenous and exogenous spatial lags as in Burridge (1981). It was a global model because it had the lagged y. Add lag X to SAR and we had the SDM.

$$Y = \rho(I_T \otimes W_N)Y + X\beta + \theta(I_T \otimes W_N)X + (\tau_T \otimes I_N)\varepsilon + (I_T \otimes \tau_N)\eta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_{NT}).$$

$$y = \rho W y + X\beta + W X\theta + \varepsilon. \qquad (2.6)$$

The k x 1 parameter vector $\theta$ measured the marginal impact of the explanatory variables from neighboring observations on the dependent variable y. Multiplying X by W produced spatial lags of the explanatory variables that reflected an average of neighboring observations X−values. Where Y was the (n×1) vector of observations on the dependent variable; X was an (n×k) matrix of observations on k exogenous variables with $\beta$ as the corresponding (k×1) vector of parameters; $\varepsilon$ was the (n×1) vector of independent and identically distributed error terms; $W$ was a (n×n) non-stochastic spatial weights matrix, with zeros on the main diagonal and non-negative values for $W_{ij}$ , i ≠j. Conventionally, W was normalized so that rows sum to 1. In this case, the endogenous lag $\sum_{j=1}^{n} W_{ij} Y_j$ was the weighted average of $Y_i$ in locations j= 1,…,n for which $W_{ij}$>0. The parameter $\rho < 1$ quantified the spatial dependence of Y on connected regions, as designated by the non-zero elements of W.

Xiao and Mao (2020) explored on postgraduate education's influence and spatial effects on technological innovation using China's provincial panel data from 2004 to 2018 based on the spatial Durbin model. The study results revealed that distribution of postgraduates in China showed spatial autocorrelation and non-equilibrium and that postgraduate education positively impacted technological innovation.

Tientao *et al.* (2016) used the SDM obtained from a theoretical model and captured the technology spillovers from a sample of 107 countries for the period 2000–2011.

### 2.3.6 Manski all-inclusive model (General Nesting Spatial Model (GNS))

Spatial models could have had three interaction effects which were correlated, endogenous or exogenous as identified by Manski (1993). All these three were contained in this model. It was given by;

$$Y = \rho(I_T \otimes W_N)Y + X\beta + \theta(I_T \otimes W_N)X + (\tau_T \otimes I_N)\mu + (I_T \otimes \tau_N)\eta + \mu.$$

$$\mu = \lambda(I_T \otimes W_N)\mu + \nu. \qquad \nu \sim N(0, \sigma^2 I_{NT}).$$

$$y = \rho Wy + X\beta + WX\theta + \mu. \qquad (2.7)$$

where $\mu = \lambda W\mu + \varepsilon$.

Polyakov *et al.* (2015) capitalized on the Manski model to estimate the private benefits of native vegetation on rural properties in the state of Victoria, Australia. Pandit *et al*. (2014) estimated the effect of tree canopy cover on sales price of urban residential properties in Perth, Western Australia using a data set of 5606 single family homes sold in 2009 and concluded that spatial effects belonged to the estimated Manski model with spatiotemporal lag.

### 2.3.7 SARAR, Cliff-Ord model or SAC (Spatial autoregressive confused) model

It was clearly defined as the spatial autoregressive model with spatial autoregressive disturbances (errors). It contained spatial lags in the dependent variable, exogenous variables, and the disturbance terms. It was given by;

$$Y = \rho(I_T \otimes W_N)Y + X\beta + (\tau_T \otimes I_N)\mu + (I_T \otimes \tau_N)\eta + \mu.$$

$$\mu = \lambda(I_T \otimes W_N)\mu + \nu. \qquad \nu \sim N(0, \sigma^2 I_{NT})$$

$$y = \rho Wy + X\beta + \mu. \qquad (2.8)$$

where $\mu = \lambda W\mu + \varepsilon$. If all $\theta = 0$,

Using the same W matrix for lag and error terms could have led to identification problems as indicated by Kelejian and Prucha (2010). Mitra and Saphores (2016) deduced on the SARAR spatial model that estimated 526 observations from a random sample collected via in-person interviews indicating that the rent of a multi-unit dwelling decreases by 0.0239% for every 1% increase in network access distance to the nearest major road in Southern Asia-Rajshahi City, Bangladesh. Olubusoye *et al.* (2016) also concluded from his study that the number of road traffic crashes in a given Local Government Area is affected by the number of road traffic crashes for neighbouring Local Government Areas in Oyo state, Nigeria based on the SARAR model.

### 2.4 Contiguity-based spatial neighbors

Apart from military benefits, Yoo and Steckel (2010) proved that mapping was critical for the institutional foundation of economic development and also for civilian and cadastral purposes. The contiguity condition was satisfied when at least one point on the border of one polygon was within the snap distance of at least one point of its neighbor and not otherwise. The contiguity relations for each county were to be recorded in the row of the matrix W. The guiding principles in the selection of ways to define the matrix W were to be according to the nature of the problem being modeled. To remove the border and boundary effects the contiguity-based spatial neighbors (queen method) was used by calculating the weight matrix (W) for the spatial data frame. $W_N$ was an N × N weight matrix, in which the elements represented the contiguity of the counties. The element on the row and column equaled 1 if the county and county had a mutual border, otherwise, it equaled 0. Corrado and Fingleton (2012) showed that when it came to the W matrix, the economic foundation of spatial models was at its weakest. Nevertheless, they concluded that the weight matrix was undeniably necessary, an important representation of spatial interaction either in the form of endogenous or exogenous lagged variable and as part of an explicit error process.

LeSage (2015) discussed the linear, rook, bishop, double linear, double rook contiguity, and queen contiguity methods. In this case, the queen method was more elaborate as it accommodated the locations that shared a common side or vertex with the county of interest.
Furthermore, this association was well reflected in the R software and was given by the argument “queen=TRUE”. When 3 or more polygons met a single point, they all met the contiguity condition giving rise to crossed links. If “queen=FALSE”, at least 2 boundary points must have been within the snap distance of each other and thus a “rook” relationship. Once the list of neighbors was made for the study area then the spatial weights were assigned. The nb2listw function in R took a neighbors list object and converted it into a weighted object. Row standardization was made where the weights for each areal fraternity were standardized to sum to unity Bivand *et al*. (2008).

## 2.5 Summary of the literature

The inception of the GCP in February 2019 had created a platform for screening the economic performance of Kenya from decentralized units. This was by way of Kenya Accountable Devolution Program (KADP) which was a World Bank initiative. This enabled further computation for utility and better comprehension of the county economies through spatial modeling. The 18 indicators of the GCP were used to divide the overall GDP as estimates of economic performance for the 47 counties. The economic survey report released in the KNBS (2019) highlighted the country's economic performance for the year 2018. The report said that the economy had grown by 6.3%, compared to 4.9% in 2017. Kwame (2019) discussed on the current GCP report in summary and gave four points that were a lesson from it. He stated that the estimates of the output from Kenya counties gave a highly valued report by policy audiences and institutions not found in the public sector. The data presentation in the report though was only through one map showing the GCP values per county unlike in this study where the autocorrelation pattern and each indicator were observed in details. The GCP 2019 report exposed the falsehood that only some specific regions of the country exclusively drove the Kenyan economy based on tribal measures which was also clearly expounded in this study. The report showed that the country was becoming urbanized and there was an added impression that the county governments were being put under implicit pressure to make use of the GCP data to design revenue policy. The council of governors was expected to see the data as necessary for the planning policy which was the core results of this study. The GCP dataset for 2019 also gave out the top county contributors to Kenya's wealth which was not sufficient and elaborate because the geographical effects were not outlined. The contribution from each county in this study was given in form of marginal effects which considered the contiguity patterns. According to Ngugi (2019), the first ever GCP 2019 study by the KNBS provided a measure of how much each county contributed to Kenya's GDP. The same concept was implemented in South Africa since 2007 and Kenya was the second country in Africa to have such a regional economic report. He stated in summary that the study tracked the monetary measure of the market value of all the final goods and services produced within each of the 47 counties.

To develop a spatial model, it was necessary to check if there was a degree of dependency among indicators in the geographic space. Unlike the descriptive summaries of data matrices through factor analysis which had been commonly interpreted in countries as for Bai *et al*. (2015) study and to the level of Pakistan districts as by Syed *et al.* (2013) and the number of factors influencing the GDP classified as in Akhter *et al*. (2012) for Bangladesh GDP and Syed *et al.* (2013), this study dealt exceptionally with Kenyan counties and went further to find out the latent variables which were used in latent spatial regression equations as in Oud and Folmer (2008).

Though both Okwi *et al*. (2007) checked on poverty in Kenya and Bai En Chong *et al.* (2012) on economic growth in China at provincial levels, this study developed spatial regression models like Afifah *et al.* (2021), Septiawan *et al.* (2018) and Karim *et al.* (2017) in Indonesia, Jiang *et al.* (2017) for China provinces, Xiao and Mao (2020) for China provinces, Tientao *et al.* (2016) for 107 countries, Polyakov *et al.* (2015) in Australia, Pandit *et al*. (2014) in western Australia, Olubusoye *et al.* (2016) in Nigeria ,Pozdílková and Marek (2018) in Czech republic, Kanaroglou *et al*. (2013) in Canada, Huang (2014) in China, Mitra and Saphores (2016) in Bangladesh and Szendi (2016) across Germany and Hungary; but for the economic performance at the county levels of Kenya. Furthermore, we checked on spill overs as in Moreno and Trehan (1997) and using the Likelihood ratio test as in Xiangyu *et al.* (2018) in China provinces to determine whether to use a restricted form of a spatial regression model.

# CHAPTER THREE
# METHODS AND MATERIALS

## 3.1 The GCP dataset

The data set which was used in this study was secondary from Kenya National Bureau of Statistics (KNBS) website. Each of the 47 counties of Kenya produced values in millions of Kenya shillings at both constant and current prices. For this study, the current price of goods and services as on 2019 dataset were used. It consisted of columns with the 47 county names and their index, shape length and area for each county, the 18 indicators and the Gross county product value for each county. The equation generated from the dataset was the relationship of the GCP values and the indicators which was assumed to be linear as shown below:

Gross County Product (GCP)= Education (**Educ**) + Agriculture, forestry and fishing (**AgrForFish**) + Mining and quarrying (**MinQua**) + Manufacturing (**Manuf**) + Water supply; waste collection (**WatSuWCol**) + Construction (**Constr**) + Wholesale and retail trade; repair of motor vehicles (**WholRRMV**) + Transport and storage (**TranSto**) + Accommodation and food service activities (**AccoFodSA**) + Information and communication (**Infocom**) + Financial and insurance activities (**FinInsuA**) + Real estate activities (**RelEstA**) + Professional, technical and support services (**ProTecSupS**) + Public administration and Defence (**PubAdminD**) + Electricity supply (**ElectSup**) + Human health and social work activities (**HumHelSocWA**)+ Other service activities (**Other Services**)+ Financial intermediation services indirectly measured (**FinIServM1**). (3.1)

It was important to note that equation (3.1) was evidently miscalculated due to large number of the predictors and thus produce inconsistent results. Furthermore, the GCP was a geographically oriented and thus the above relationship produces absurd results.

### 3.1.1 The study area

The study area was the country Kenya which had 47 counties. It was found in East Africa and apart from the counties which define geographical boundaries, there existed 6 economic blocks which included Frontier Counties Development Council (FCDC) (7 counties), North Rift Economic block (NOREB) (8 counties), Lake Region economic block (LREB) (14 Counties), Jumuia ya kaunti za pwani (6 counties), South Eastern Kenya Economic Block (3 counties), Mt.Kenya and Aberdares Region Economic block (10counties).

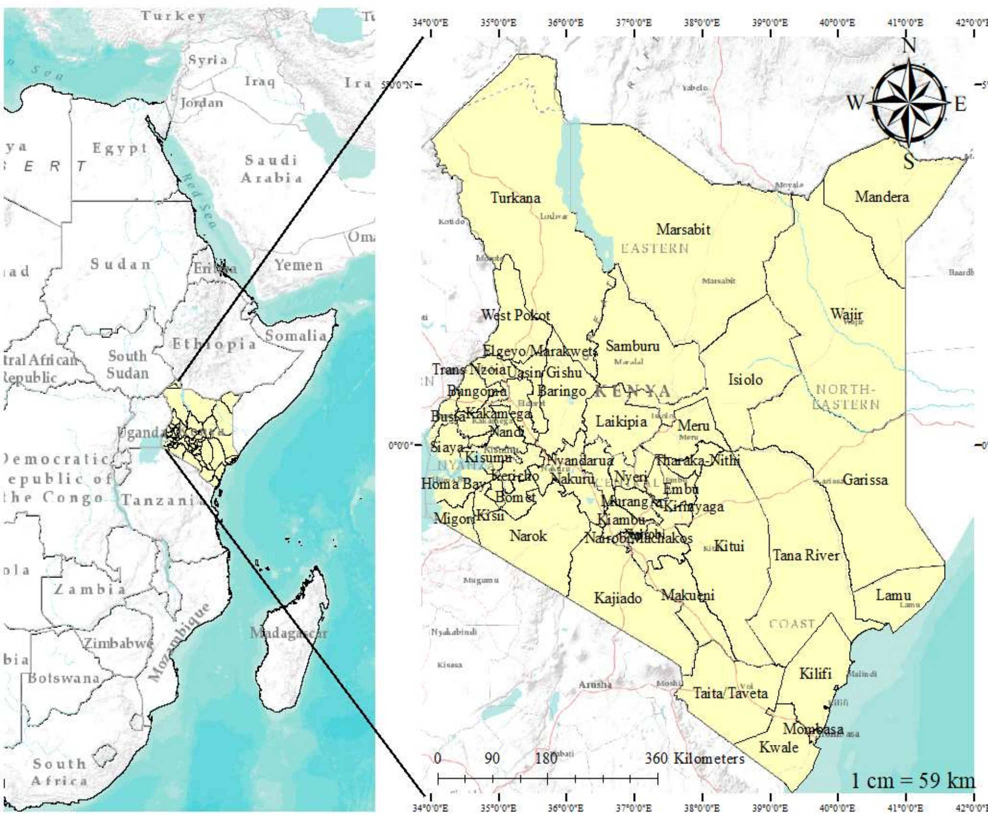



**Figure 3.1:** The study area

### 3.2 Factor analysis of GCP data

The analysis simplified complex and diverse relationships that existed among the indicators by revealing common dimensions or factors that connected the seemingly unrelated variables and consequently provided insight into the importance of the underlying structure of the data. Factor analysis was used for data reduction to identify the small number of factors that discussed most of the variances observed in the much larger number of manifest variables. The underlying assumption of the analysis is that there existed some independent variable (latent variables) that accounted for the correlations among dependent variables all becoming zero. In other words, the latent variables determined the values of the dependent variables.

The dataset was checked for sampling adequacy, reliability, and sphericity as a necessity of the factor analysis model through the Kasier-Meiyer-Olkin (KMO) test, and Cronbach and Bartlett's test respectively. Factor analysis could have been done in different algorithm forms which included minimum residual method (MRM), principal axes, alpha factoring, weighted least squares, minimum rank, or maximum likelihood method (MLM). The minimum residual (MRM) solution was an unweighted least squares solution that took a slightly different approach. Principal axes (PA) could have been used in cases when maximum likelihood solutions failed to converge. MRM also worked alternatively tending to produce better and smaller Root mean square error (RMSEA) of approximation solutions. The maximum likelihood solution found those communality values that minimized the chi-square goodness of fit test producing a more expansive output. The maximum likelihood factor analysis was probably preferred as suggested by Revelle (2019).

In comma-separated values (.csv) format, the data was loaded into R software for factor analysis and the Maximum Likelihood Method (MLM) employed. The factor score output that represented each county was bound together with the original dataset by dropping the duplicate columns. Then the newly formed set now consisted of the county names, the factor scores, geographical shapes, and GCP values in millions of Kenyan shillings. The set was converted to shape file (.shp) format in ArcGIS and used for geospatial mapping and modeling the dependence of the GCP values on the latent variables which were the factor score columns.

### 3.3 Exploratory data analysis

A summary of the indicator variables which consisted of the range, mean, median, kurtosis, and skewness and Inter Quartile range (IQR) values was obtained. Also their Pearson's Correlation values were presented graphically using colors with the intensity revealing strength of their relationship. Thematic maps that present each of the indicators graphically were plotted and clearly labelled.

### 3.4 Spatial dependence

#### 3.4.1 Spatial autocorrelation

The prior obligation in this study was on finding the presence of spatial autocorrelation where variables were correlated by themselves based on a measure of the systematic pattern in the dispersion of objects within a space. Spatial autocorrelation therefore suggested that observations at a location depended on observations in other locations that shared similar characteristics as suggested by Anselin and Arribas-Bel (2013).

The response variable in the spatial econometric model was the GCP of which was assumed to have a spatial dependence on the latent values. The Local Indicator of Spatial association (LISA) principle (Moran I test) was used to find out the presence of spatial autocorrelation from the simple linear model and the Morans' I scatterplot as in Páez *et al.* (2009). A map of whether the relationship was positive or negative was displayed for clear reference in the final report. The hypothesis to be used was:

$\mathrm{H_0{:}I_M} = 0$ (no autocorrelation)

$\mathrm{H_1{:}I_M} \neq 0$ (There is autocorrelation)

The test statistic was as follows:

$$Z_h = \frac{I_M - I_{M_O}}{\sqrt{Var(I_M)}}. \tag{3.2}$$

where, $I_M = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ and $Var(I_M) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)S_0^2}$.

The p-value of the Moran I statistic gave the direction to proceed to the next step on the source of the spatial effects.

### 3.4.2 Determining the weight matrix

The weight matrix W was formed from the queen method of contiguity that considered shared vertices and the number of non-zero links found after row standardization. The spatial weights object formed was then checked if it was similar to symmetry and also transformed as a sparse matrix to yield real eigenvalues or for cholesky decomposition.

## 3.5 Fitting the spatial econometric models

### 3.5.1 Parameter estimation using the maximum likelihood estimation method

Spatial modeling included reporting on the parameter derivations below using the method of maximum likelihood estimation which stated that if we had a random sample from the probability density function $f(x_i;\theta)$ and we were interested in estimating θ, the maximum likelihood estimator denoted $\hat{\theta}_{\text{mle}}$ was the value of θ that maximized$L(\theta|x)$. With random sampling, the log-likelihood had the particularly simple form below as in Zivot (2009).

$$\ln L(\theta|x) = \ln(\prod_{i=1}^{n} f(x_i;\theta)) = \sum_{i=1}^{n} ln\, f(x_i;\theta). \tag{3.3}$$

### 3.5.2 Parameter estimation for the simple linear model

The ordinary least squares regression model was given by equation (2.1). Using the Maximum likelihood technique,

$\underline{Y} \sim N_n(X\beta, \sigma^2 \text{I})$ and $\underline{\text{e}} \sim N_n(\underline{O}, \sigma^2 \text{I})$

With the normality assumption, the MLE was obtained by denoting the likelihood function:

$L(\underline{\beta}, \sigma^2)$. Hence the values of $\beta_i$ and $\sigma^2$ maximized $L\left(\underline{\beta}, \sigma^2\right)$.

$L\left(\underline{\beta}, \sigma^2\right) = f(\underline{y}, \underline{\beta}, \sigma^2)$.

$$= \frac{1}{(2\pi)^{n/2} |\sigma^2 \text{I}|^{1/2}} \exp(-\frac{1}{2}((Y - X\beta)' \sigma \text{I}(\text{Y} - \text{X}\beta)). \tag{3.4}$$

Therefore, the parameters obtained were $\beta_i$ and $\sigma^2$ which were reported in the results.

### 3.5.3 Parameter estimation for the Spatially Lagged X (SLX) model

This was given by equation (2.2). The parameter estimates by Maximum likelihood estimation were as follows:

$L\left(\underline{\beta}, \sigma^2\right) = f(\underline{y}, \underline{\beta}, \sigma^2)$.

$$= \frac{1}{(2\pi)^{n/2}|\sigma^2 I|^{1/2}} \exp\left(-\frac{1}{2}(Y - X\beta - WX\theta)'\sigma I(Y - X\beta - WX\theta)\right) \quad (3.5)$$

The parameters to be obtained were $\beta_i$, $\sigma^2$ and $\theta$ values which was a k x 1 parameter vector.

**3.5.4 Parameter estimation for the Spatial Lag Y (Autoregressive) model (SAR)**

This was given by equation (2.3). Using the maximum likelihood function the parameters of the SAR model were estimated as follows.

$$\ln L(\underline{\beta}, \sigma^2) = -{}^{n}/_{2} \ln(2\pi) - {}^{n}/_{2} \ln(\sigma^2) - {}^{1}/_{2\sigma^2} (Y - X\beta - \rho WY)'(Y - X\beta - \rho WY) \quad (3.6)$$

The computationally troublesome aspect of this was the need to compute the log-determinant of the $n \times n$ matrix $(I_n - \rho W)$. Nevertheless, we had $\beta_i$, $\sigma^2$ and $\rho < 1$ parameter value which quantified the spatial dependence of Y on connected regions.

**3.5.5 Parameter estimation for the Spatial Error Model (SEM)**

This was given by equation (2.4). The SEM model had a concentrated log-likelihood taking the form:

$$lnL = C + \ln|I_n - \lambda W| - (n/2)\ln(\varepsilon'\varepsilon).$$

In details we had;

$$\mu \sim N(\underline{O}, \sigma^2(I - \lambda W)'(I - \lambda W)).$$

Let $(I - \lambda W) = Z$ and

$$lnL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(Z'Z) - \frac{1}{2}(Y - X\beta)'(Z'Z)^{-1}. \quad (3.7)$$

Thus, we obtained $\beta_i$, $\sigma^2$ and $\lambda$ which was the spatial dependence parameter.

**3.5.6 Parameter estimation for the Spatial Durbin Error Model (SDEM)**

This was given by equation (2.5).The error term $\varepsilon$ was expressed as a function of the vector having auto correlated disturbances $\mu$.

$$\varepsilon = \mu(I - \lambda W) \text{ and } \mu = (I - \lambda W)^{-1}\varepsilon.$$

Therefore equation (2.5) became

$$Y = X\beta + WX\theta + (I - \lambda W)^{-1}\varepsilon. \quad (3.8)$$

Let $Y - WX\theta = Z$ and $I - \lambda W = A$ thus, $(Z - X\beta)'A = \varepsilon$.

Substituting in the Log-likelihood function we had;

$$(X'A'AZ)(X'A'AX)^{-1} = \hat{\beta}.$$

$$[X'(I - \lambda W)'(I - \lambda W)(Y - WX\theta)][X'(I - \lambda W)'(I - \lambda W)X]^{-1} = \hat{\beta}. \quad (3.9)$$

The parameters obtained were $\beta_i, \sigma^2$ ,$\lambda$ which was the spatial dependence parameter and $\theta$ values which was a k x 1 parameter vector.

### 3.5.7 Parameter estimation for the Spatial Durbin Model (SDM)

This was given by equation (2.6). The log likelihood form took the following format;

$$\ln L(\underline{\beta}, \sigma^2) = -{}^{n}/_{2} \ln(2\pi) - {}^{n}/_{2} \ln(\sigma^2) - {}^{1}/_{2\sigma^2} (Y - X\beta - \rho WY - WX\theta)'(Y - X\beta - \rho WY - WX\theta). \quad (3.10)$$

The parameters were $\beta_i, \sigma^2$, $\rho < 1$ parameter value which quantified the spatial dependence of Y on connected regions and $\theta$ values which was a k x 1 parameter vector.

### 3.5.8 Parameter estimation for the Manski All-inclusive Model (GNS model)

This was given by equation (2.7). From equation (2.7) we had

$$y = \rho Wy + X\beta + WX\theta + (I - \lambda W)^{-1}\varepsilon. \quad (3.11)$$

and $(y(I - \rho W) - X\beta - WX\theta)'(I - \lambda W) = \varepsilon$

Let $[y(I - \rho W) - WX\theta] = Z$ and $(I - \lambda W) = A$

Substituting in the log-likelihood function;

$$(X'A'AZ)(X'A'AX)^{-1} = \hat{\beta}.$$

$$[X'(I - \lambda W)'(I - \lambda W)(Y(I - \rho W) - WX\theta)]^{-1} = \hat{\beta}. \quad (3.12)$$

The parameters were $\beta_i, \sigma^2, \lambda$ which was the spatial dependence parameter, $\rho < 1$ parameter value which quantified the spatial dependence of Y on connected regions and $\theta$ values which was a k x 1 parameter vector.

### 3.5.9 SARAR, Cliff-Ord model or SAC (Spatial Autoregressive Confused) model

This was given by equation (2.8). From equation (2.8) we had

$$Y = \rho WY + X\beta + (I - \lambda W)^{-1}\varepsilon. \qquad (3.13)$$

and $(Y - \rho WY - X\beta)'(I - \lambda W) = \varepsilon$.

Let $Y(I - \rho W) = Z$ and $(I - \lambda W) = A$.

Substituting in the Log-likelihood model;

$(X'A'AZ)(X'A'AX)^{-1} = \hat{\beta}$.

$$X'(I - \lambda W)'(I - \lambda W)(I - \rho W)Y][X'(I - \lambda W)'(I - \lambda W)X]^{-1} = \hat{\beta} \qquad (3.14)$$

The parameters were $\beta_i, \sigma^2, \lambda$ which was the spatial dependence parameter and $\rho <$ 1 parameter value which quantified the spatial dependence of Y on connected regions.

### 3.6 Model selection and comparison

Presence of spatial autocorrelation prompted for a Lagrange Multiplier/Rao score test (LM/RS test) to check whether the spatial effects were displayed significantly in a residual pattern or through lags. The Anselin (1988) method was used to decide which spatial model was to be preferred to satisfy underlying objectives. The Anselin method suggested that if only one of the LMerr (Lagrange Multiplier Error test) and LMlag (Lagrange Multiplier Lag test) were significant, then an extra step was taken by checking the robust (false positives for the other kind of spatial relationship) versions RLMerr (Robust Lagrange Multiplier Error test) and RLMlag (Robust Lagrange Multiplier Lag test). If only one of them was significant, then that model was adopted and SARMA (Spatial Autoregressive Moving Average) model had to be ignored. An original suggestion of the LM/RS test against a spatial error alternative was made by Burridge (1981) and took the form.

$$LM_{err} = [\varepsilon'W\varepsilon/(\varepsilon'\varepsilon/N)]^2/[tr(W^2 + W'W)].$$

The $LM_{err}$had an asymptotic $\chi^2(1)$ distribution similar to the LM/RS test against a spatial lag alternative which was given by Anselin (1988) as

$$LM_{lag} = [\varepsilon'Wy/(\varepsilon'\varepsilon/N)]^2/D$$

where $D = [(WX\beta)'(I - X(X'X)^{-1}X')(WX\beta)/\sigma^2] + tr(W^2 + W'W)$.

The SDEM, SDM, Manski and SARAR could be nested or restricted back to simpler models after testing positive on lack of fit using the Likelihood ratio test which was a maximum likelihood based specification test. A pairwise selection was conducted from the nested models to the simple linear ones which provided a significant explanation of the spatial phenomenon. Likelihood ratio test was the difference between the log-likelihoods of the nested models and was given by

$$\lambda_{LR} = -2[l(\theta_0) - l(\hat{\theta})]. \tag{3.15}$$

Based on the ratio of the likelihoods of any two participating models that are nested, the one with the best fit was acquired as in King (1989).

For example, if the null hypothesis of the LR test (H0: $\gamma + \delta\beta = 0$) could not be rejected, then the Manski model (2.7) which was nested could be simplified to the SLX (2.2) as in Burridge (1981) i.e. from equation (3.15), we had

$$\lambda_{LR} = -2[l(\theta_7) - l(\theta_2)].$$

Restriction of the spatial econometric models to the simple linear model was as follows according to Burkey (2018);

i. The Manski all-inclusive model to the SARAR , the Spatial Error model (SEM), the Spatial Lag Y(Autoregressive) Model (SAR) and finally the simple linear model or
ii. The Manski all-inclusive model to the Spatial Durbin Model (SDM), the Spatially Lagged X (SLX), the Spatial Lag Y (Autoregressive) model (SAR), the Spatial Error model (SEM) and finally the simple linear model.
iii. The Manski all-inclusive model to the Spatial Durbin Error model (SDEM), the Spatial Error model (SEM), the Spatially Lagged X (SLX) and finally the simple linear model.

The spatial Hausman test was used to compare the parameter estimates of the simple linear model and the SEM. If there was statistical significance in the test then it could have meant that neither of the models was fit for the data. There could have been spatial dependence but the SEM was not the appropriate model to capture the spatial phenomenon as suggested by LeSage and Pace (2009).

The methods used in checking the corresponding standard of the models included the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC was used to estimate the quality of a model relative to another for the same data. It was given by;

$$AIC = 2k - 2\ln(\hat{L}). \tag{3.16}$$

where k was the number of estimated parameters in the model and $\hat{L}$ the maximum value of the likelihood function for the model. The AIC value for each model was obtained and the best model was the one with the minimum value as in Akaike (1974).

The BIC was a criterion similar to AIC but had a larger penalty term and was given by;

$$BIC = kln(n) - 2\ln(\hat{L}). \tag{3.17}$$

where k was the number of estimated parameters in the model, $\hat{L}$ the maximum value of the likelihood function for the model and n the number of observations. The BIC was evaluated for each of the models and the model respective to the minimum value of the BIC was selected as in Schwarz (1978).

**3.7 Data analysis**

The statistical packages that were used to analyse the data was ArcGIS (version 10.7.1) and R (version 4.3.0). In ARCGIS, the shapefile that consisted of the spatial data frame was created and then loaded into R software where the spdep, leaflet, tmap, raster, sf, rgeos, rgdal and spatialreg packages brought forth the results of the spatial relationship and the tests mentioned in this chapter. The R software gave a clear platform for spatial modeling and raster image creation. For factor analysis, the factor scores from the n-factor model that had the lowest BIC were the ones going to be used for creating latent regression equations. The Figure 3.2 below summarized the whole analysis process that was implemented in this study.

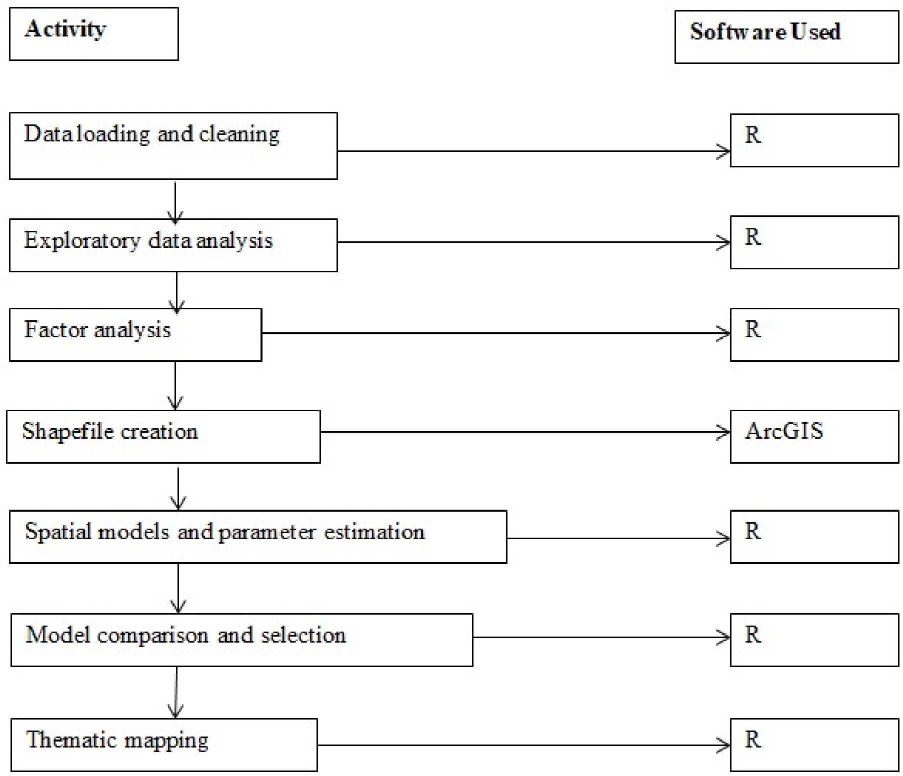



**Figure 3.2:** Data analysis guideline

## CHAPTER FOUR
## RESULTS AND DISCUSSION

### 4.1 Introduction to results of the study

The chapter consists of the results from exploratory data analysis, exploratory factor analysis, spatial autocorrelation (Moran I test), simple linear model (OLS), and spatial regression with model comparison tests and thematic maps. The summary of the variables that were used in the study were presented in the following table showing the measures of central tendency and dispersion in millions of Kenyan shillings together with their skewness and kurtosis:

**Table 4.1:** GCP indicator variables summary in millions of Kenyan Shillings

| Indicator | Min | 1st Q. | Median | Mean | 3rd Q. | Max | Skew | Kurt. | IQR |
|---|---|---|---|---|---|---|---|---|---|
| Agriculture, forestry and fishing | 1459.00 | 19945.00 | 47606.00 | 60404.00 | 78382.00 | 301349.00 | 2.15 | 6.04 | 58436.50 |
| Electricity supply | 22.00 | 346.00 | 581.00 | 2994.00 | 1310.00 | 36932.00 | 3.38 | 11.27 | 964.00 |
| Mining and quarrying | 40.00 | 220.00 | 620.00 | 1244.00 | 1444.00 | 9643.00 | 2.84 | 9.57 | 1224.50 |
| Manufacturing | 11.00 | 119.00 | 1153.00 | 13769.00 | 6018.00 | 374527.00 | 6.02 | 36.48 | 5899.50 |
| Water supply; waste collection | 90.00 | 441.00 | 740.00 | 1191.00 | 1142.00 | 10819.00 | 3.97 | 16.91 | 700.50 |
| Construction | 24.00 | 1604.00 | 3184.00 | 9626.00 | 6386.00 | 175437.00 | 5.34 | 29.88 | 4783.00 |
| Wholesale and retail trade; repair of motor vehicles | 1257.00 | 3542.00 | 5051.00 | 13186.00 | 7372.00 | 294302.00 | 6.18 | 37.96 | 3830.00 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Transport and storage | 258.00 | 3260.00 | 5708.00 | 12771.00 | 10076.00 | 184845.00 | 4.80 | 24.33 | 6815.00 |
| Accommodation and food service activities | 45.00 | 166.00 | 337.00 | 1237.00 | 811.00 | 14041.00 | 3.58 | 12.29 | 645.00 |
| Information and communication | 143.00 | 427.00 | 881.00 | 2329.00 | 1684.00 | 53074.00 | 6.19 | 38.08 | 1257.50 |
| Financial and insurance activities | 260.00 | 3876.00 | 7380.00 | 12897.00 | 15414.00 | 142765.00 | 4.82 | 26.18 | 11537.00 |
| Real estate activities | 752.00 | 3090.00 | 5733.00 | 12242.00 | 10029.00 | 176281.00 | 5.35 | 30.58 | 6939.50 |
| Professional, technical and support services | 1.00 | 5.50 | 21.00 | 2920.30 | 228.50 | 122335.00 | 6.39 | 39.82 | 223.00 |
| Public administration and Defence | 2129.00 | 4756.00 | 5973.00 | 7033.00 | 7282.00 | 40051.00 | 4.42 | 23.01 | 2525.50 |
| Education | 923.00 | 4114.00 | 6252.00 | 6813.00 | 9348.00 | 16676.00 | 0.58 | -0.33 | 5234.00 |
| Human health and social work activities | 254.00 | 1226.00 | 2248.00 | 2696.00 | 3148.00 | 17841.00 | 3.59 | 16.17 | 1923.00 |
| Other Variables | 257.00 | 1138.00 | 1839.00 | 1952.00 | 2280.00 | 8791.00 | 2.65 | 10.56 | 1141.00 |
| Financial Intermediation Services Indirectly | 78.00 | 375.50 | 709.00 | 5204.60 | 1497.50 | 168283.00 | 6.82 | 38.88 | 1122.00 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Measured | | | | | | | | | |
| Total GCP | 15850.00 | 72997.00 | 103734.00 | 160100.00 | 173990.00 | 1492323.00 | 4.70 | 24.97 | 100993.00 |

The standard normal distribution had a kurtosis of 3 and thus the variables were largely non-normal. The Total GCP had a mean of 160100 × 10^6 and a median value of 103734× 10^6. The minimum and maximum value of the Total GCP was 15850 × 10^6 and 1492323× 10^6 respectively. It also had an IQR of 100993 × 10^6 with skewness and kurtosis at 4.7 and 24.97 respectively. Also, the first quartile value for the Total GCP is 72997× 10^6 and third quartile value is 173990× 10^6 showing the tail of less values on the right. The rest of the variables could be discussed in a similar manner.

### 4.1.1 The Correlation plot

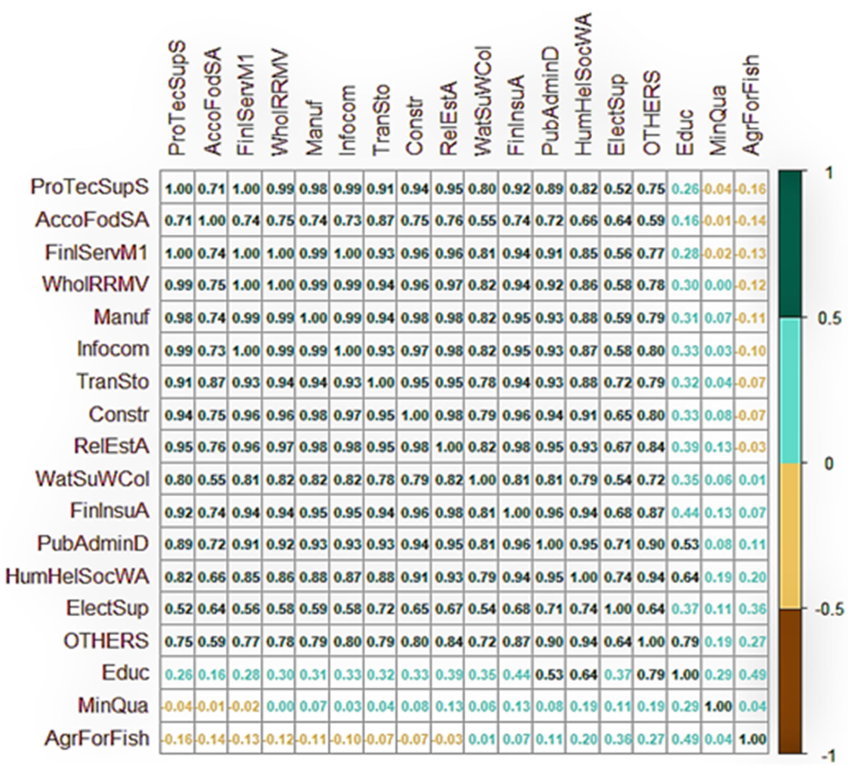



**Figure 4.1:** Pearson's correlation plot

The correlation plot in Figure 4.1 showed most of the values having a strong positive Pearson's correlation value as presented by the dark green shaded color and few weakly negative correlated values. Variables such as Education, Agriculture Forestry and fishing, Mining and Quarrying are poorly related with the rest. This large number of correlated values that was unseen in the KNBS (2019) report contributed to the ill-defined model and thus the exploratory factor analysis was performed as displayed in the next sub-topic.

### 4.1.2 Dimension Reduction Procedure: Factor Analysis

First, the reliability of the data set was checked using the Cronbach coefficient α. The reliability value of the data was **0.8861494** which was compared with the standard value alpha of 0.7 advocated by Cronbach (1951), a more accurate recommendation Nunnally and Bernstein's (1994) or with the standard value of 0.6 as recommended by Bagozzi and Yi's (1988) was found out to be that the scales used for the secondary data were sufficiently reliable for data analysis.

To make sure that the dataset was factorable the Bartlett's test of Sphericity and the Kasier – Meyer –Olkin (KMO) validity test were conducted where in the KMO test, a measure greater than the value of 0.8 is considered meritorious. The data returned a value for the overall sampling adequacy of **0.89** indicating meritorious. Bartlett's test of Sphericity in Figure 4.2 was a measure of the multivariate normality of the set of distributions. It also tested whether the correlation matrix conducted within the FA was an identity matrix. FA would have been meaningless with an identity matrix. A significance value less than 0.05 indicated that the data did not produce an identity matrix and was thus appropriately multivariate normal and acceptable for FA as by George and Mallery (2021). The data within this study returned a significance value of **1.74**, indicating that the data was acceptable for FA.

```
$chisq
[1] 545.5772

$p.value
[1] 1.737363e-45

$df
[1] 153
```

**Figure 4.2:** Bartlett's test of Sphericity

The scree plot below in Figure 4.3 showed the number of factors for the exploratory factor analysis (EFA) but it was not possible to determine the number of factors to be used in the factor analysis from it.

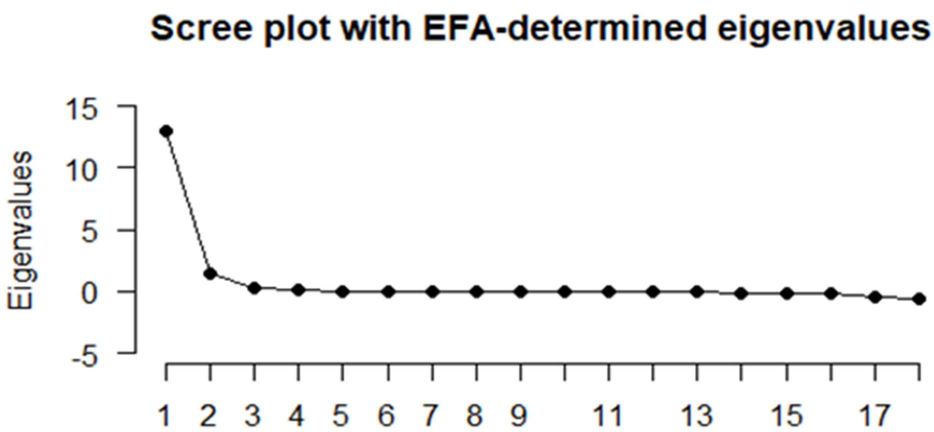



**Figure 4.3:** Exploratory Factor Analysis Scree plot

The factor analysis models for each number of factors were compared for best fit using the Bayesian Information Criterion (BIC) in an ANOVA and the result was 4 factors as shown below in Table 4.2.

**Table 4.2:** Factor Analysis Model Comparison

| Factor Analysis Model | BIC |
|---|---|
| 1st Model | -67.54 |
| 2nd Model | -148.39 |
| 3rd Model | -144.12 |
| 4th Model | **-159.32** |
| 5th Model | -114.04 |

Model 4 had the deepest negative BIC = **-159.32** and thus was suitable for the EFA.

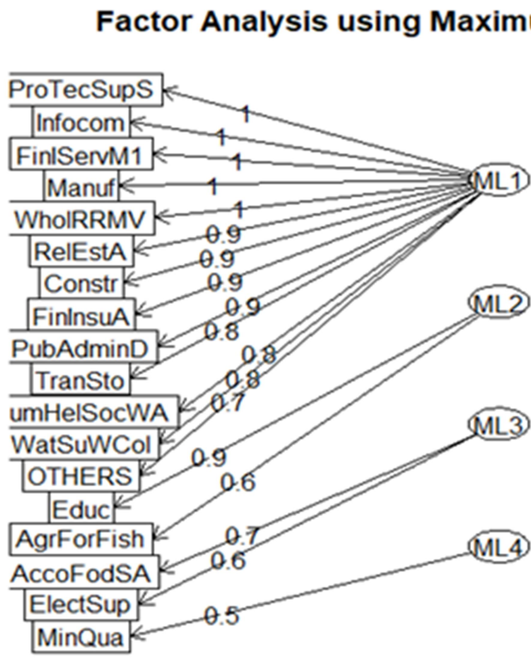



**Figure 4.4:** Factor analysis Path diagram

Based on the path diagram in Figure 4.4 which consisted of the factor loadings and the indicators, it appeared that 4 components existed. With a strong correlation between the factors and the indicator variables, a factor analysis with 4 factors was conducted. From the factor analysis, the first 4 factors explained almost **87.32%** of the total variance. The Maximum likelihood method of factor analysis was used. The path diagram revealed that there were four groupings that formed up that showed some affiliation in the mapping.

From the analysis, it was evident that four major factors were influencing Kenya's GCP. The first factor consisted of Professional, technical and support services (ProTecSupS), Financial intermediation services indirectly measured (FinIServM1), Information and communication (Infocom), Wholesale and retail trade; repair of motor vehicles (WholRRMV), Manufacturing (Manuf), Real estate activities (RelEstA), Construction (Constr), Water supply; waste collection (WatSuWCol), Financial and insurance activities (FinInsuA), Public administration and Defence (PubAdminD), Human health and social work activities (HumHelSocWA), Transport and Storage (TranSto), Other service activities (Other Services). Second factor consisted of Education (Educ) with Agriculture, forestry and fishing (AgrForFish) and the third factor consisted of Accommodation and food service activities (AccoFodSA) with Electricity Supply (ElectSup) and the fourth Mining and quarrying

(MinQua). The factors formed were just as seen in the general conclusion of the correlation plot before.

### 4.1.3 Spatial boundaries and weights

A 47 by 47 sparse weight matrix W was formed from the queen method of contiguity that considered shared vertices and the number of non-zero links that were found after row standardization was 232.

Weights (W) were assigned through a coding scheme style called row standardization or “w” style which sums over all links to the total number of regions. The resulting spatial weight list was checked whether it was similar to symmetry and could yield real Eigen values and also for cholesky decomposition which was verified to be true.

The use of row-standardization led to asymmetry even if the underlying neighbors were symmetric; unless all entities had matching numbers of neighbors which was shown that they were not after being plotted against randomly generated numbers from a uniform distribution in Figure 4.5.

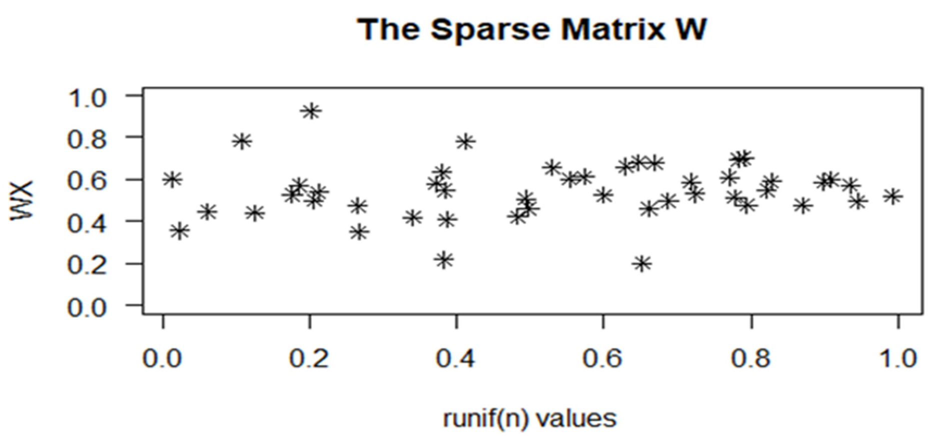



**Figure 4.5:** Sparse Matrix W Unmatched Neighbors

### 4.1.4 The simple linear model result

**Table 4.3:** OLS Results

| Coefficients | Estimate( $\beta_i$ ) | Std. Error( $\sigma^2$ ) | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 160100.00 | 6544.00 | 24.47 | 0.00 *** |
| ML1 | 193867.00 | 7206.00 | 26.9 | 0.00 *** |
| ML2 | 30982.00 | 7627.00 | 4.06 | 0.00 *** |
| ML3 | 38066.00 | 12014.00 | 3.17 | 0.00 ** |
| ML4 | -13477.00 | 11485.00 | -1.17 | 0.25 |

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**Residuals**

| Min | 1Q | Median | 3Q | Max |
|---|---|---|---|---|
| -60350.00 | -19108.00 | -5805.00 | 7901.00 | 158545.00 |

Residual standard error: 44860.00 on 42 Degrees of freedom

Multiple R-squared: 0.96 Adjusted R-squared: **0.96**

F-statistic: 270.30 on 4 and 42 Degrees of Freedom p-value: **0.00**

The OLS in Table 4.3 had a significant parameter results though the ML4 was not needed in the model. The adjusted R-squared value was **0.96** and the model had a p-value of **0.00.** According to equation (2.1), the model therefore was given as;

TOTAL = 160100+193867 ML1+30982 ML2+38066 ML3-13477 ML4

And it lacked spatial or temporal effects which were suspicious since the dataset was geographically tabulated. It was important to note also that the values in the model were in millions of Kenyan shillings and the extra zeroes were truncated.

### 4.2 Moran I test

We checked if there was any spatial autocorrelation in the model using the Moran's I test (Correlation test on the residuals).

$H_0$:$I_M = 0$ ($no\ spatial\ autocorrelation$)versus

$H_1$:$I_M \neq 0$ ($There\ is\ spatial\ autocorrelation$)

**Table 4.4:** Global Moran I Test

| | | | |
|---|---|---|---|
| Moran I statistic standard deviate = 2.8221 | | | p-value = **0.0024** |
| alternative hypothesis: greater | | | |
| Sample estimates | **Observed Moran I** | **Expectation** | **Variance** |
| | **0.2118** | -0.0391 | 0.0079 |

Conclusion: We rejected the Null hypothesis and concluded that there was **spatial autocorrelation** in the residuals and so there was a sense of spatial dependence or clustering.

The spatial matrix from the queen method listw1 was used and spatial correlation was checked in the residual values of reg1. The Moran I statistic in Table 4.4 was positive **0.2118** as from equation (3.2) and the p-value = **0.0024** as displayed below in the Moran plot for the residuals.

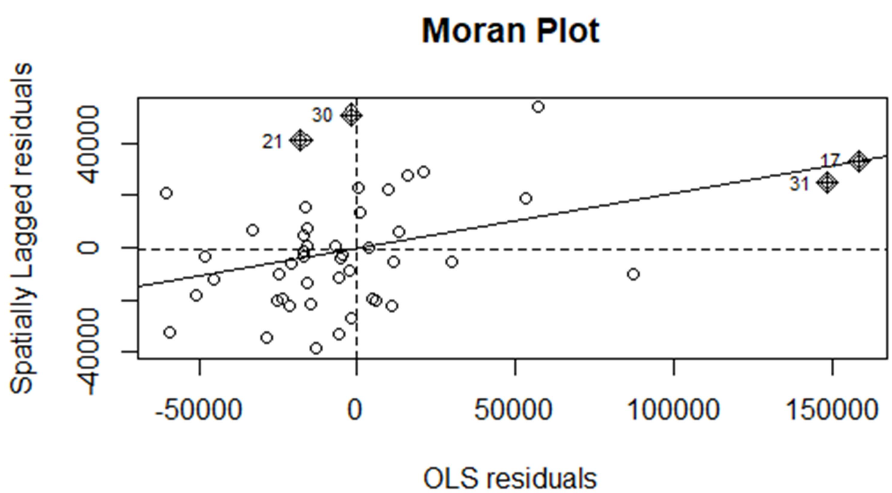



**Figure 4.6:** Moran I plot

The Moran scatterplot revealed a level of spatial association among the values and four outliers that indicated instability of the association.

### 4.2.1 Plotting the Local Moran Statistic

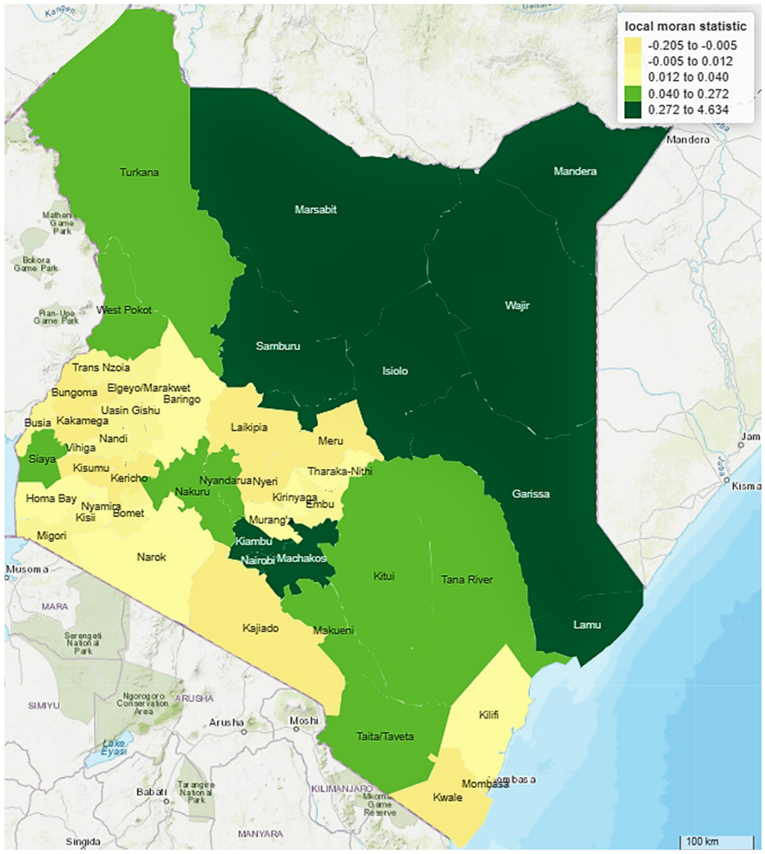



**Figure 4.7:** Local Moran Map

From Figure 4.7, it was possible to observe the variations in autocorrelation across space through the Local Indicator of Spatial association (LISA) clustering principle that was suggested by Anselin and Florax (1995). In the local Moran I statistic, each point each location received its own I value unlike the global Moran I statistic which we got **0.2118**. The interpretation was that there seemed to be a geographic pattern to the autocorrelation. However, it was not possible to understand if these were clusters of high or low values. To understand these, a map which labels the features based on the types of relationships they share with their neighbors was created (i.e. high and high, low and low, insignificant, low and high, high and low).

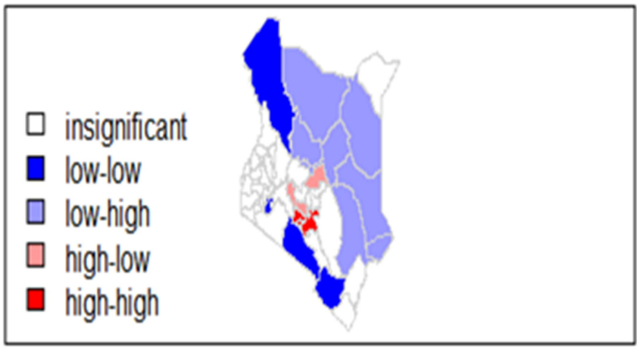



**Figure 4.8:** Total GCP Geographical Clusters

From Figure 4.8, it is apparently clear that there was a statistically significant geographic pattern to the clustering of the Total GCP in Kenya.

## 4.3 Model Selection

### 4.3.1 Lagrange Multiplier Test

The Lagrange Multiplier test was conducted to find out which spatial model could have produced a better fit to display the spatial dependence among the residuals. The Luc Anselin method decided which spatial model was to be preferred against the rest.

**Table 4.5:** Lagrange Multiplier/ Rao score Test

| Statistical Test | Statistic | Parameter | P-value |
|---|---|---|---|
| LMerr | 4.6842 | 1 | **0.0304** * |
| LMlag | 3.6417 | 1 | 0.0564 |
| RLMerr | 1.8199 | 1 | 0.1773 |
| RLMlag | 0.7774 | 1 | 0.3779 |
| SARMA | 5.4617 | 2 | 0.0652 |

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

There was no need to continue to the robust forms of the models since in Table 4.5 the LMerr was statistically significant with a p-value **0.0304** * at an alpha level of 0.05. Thus a spatial error model (SEM) was the one that was suitable for the fitting.

### 4.3.2 The Spatial Error model result

**Table 4.6:** Spatial Error Model Results

| Coefficients | Estimate( $\beta_i$ ) | Std. Error( $\sigma^2$ ) | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 159809.20 | 11882.50 | 13.45 | 0.00* |
| ML1 | 192073.20 | 6940.70 | 27.67 | 0.00* |
| ML2 | 30559.60 | 7219.70 | 4.23 | 0.00* |
| ML3 | 38472.30 | 10265.10 | 3.75 | 0.00* |
| ML4 | -22577.30 | 10562.10 | -2.14 | 0.03* |
| | | **Residuals** | | |
| **Min** | **1Q** | **Median** | **3Q** | **Max** |
| -61827.20 | -18535.00 | -5476.40 | 10087.60 | 141932.90 |
| Lambda ( $\lambda$ ):0.53 | LR test value: 6.06 | Asymptotic standard error: 0.15 | | p-value: **0.01** |
| z-value: 3.47 | | p-value: 0.00 | | |
| Wald statistic: 12.02 | | p-value: 0.00 | | |
| Log likelihood: -564.45 for error model | ML residual variance (sigma squared): 1474500000.00(sigma: 38399.00) | Number of parameters estimated: **7** | | AIC: 1142.90 |

The SEM had a significant result and all latent variables were needed in the model. The model had a p-value of **0.01**, a lambda value of **0.53** and an error term $\varepsilon$ of **38399.00.** According to equation (2.4), the model therefore was given as;

$$\text{TOTAL} = 159809.2+192073.2\ \text{ML1}+30559.6\ \text{ML2}+38472.3\ \text{ML3}-22577.3\text{ML4}+(I_{47} - 0.53W)^{-1}38399$$

where W was the weight matrix. As common, the SEM had no marginal effects and was the model that had the best fit. The vector $\mu$ for auto-correlated disturbances was given as follows;

$$\mu = 0.53W\mu + 38399.$$

$$\mu(I_{47} - 0.53W) = 38399.$$

$$\mu = (I_{47} - 0.53W)^{-1}38399.$$

## 4.4 Model comparison

### 4.4.1 Likelihood Ratio test

The table below showed the p-values of the test at an alpha level of 0.05.

**Table 4.7:** Likelihood Ratio Test

| Model | OLS | SLX | SAR | SEM | SDM | SDEM | SARAR | MANSKI |
|---|---|---|---|---|---|---|---|---|
| **OLS** | N/A | 0.0625 | 0.0477 | 0.0139 | 0.0363 | 0.0446 | 0.0268 | 0.0642 |
| **SLX** | 0.0625 | N/A | 0.1701 | **0.4095** | 0.0860 | 0.1194 | 0.4263 | 0.2276 |
| **SAR** | 0.0477 | 0.1701 | N/A | N/A | 0.0927 | 0.1140 | 0.0685 | 0.1571 |
| **SEM** | **0.0139** | **0.4095** | N/A | N/A | **0.2119** | **0.2568** | **0.2771** | **0.3214** |
| **SDM** | 0.0363 | 0.0860 | 0.0927 | 0.2119 | N/A | N/A | 0.1991 | 0.9078 |
| **SDEM** | 0.0446 | 0.1194 | 0.1140 | 0.2568 | N/A | N/A | 0.2477 | N/A |
| **SARAR** | 0.0268 | 0.4263 | 0.0685 | 0.2771 | 0.1991 | 0.2477 | N/A | 0.3233 |
| **MANSKI** | 0.0642 | 0.2276 | 0.1571 | 0.3214 | 0.9078 | N/A | 0.3233 | N/A |

From the Table 4.7 above, we did not need to restrict the error model to a lagged model as shown by the p-values of the test. It also suggested that the error model could be restricted to the OLS and all the other complex models should be restricted to the error model since there was no significance in the test. Though the OLS model produced a significant result, the criterions suggested it not to be considered as seen in Table 4.8. Further tests had to prove that the SEM model was better through the AIC/BIC.

### 4.4.2 AIC and BIC

Comparing with the rest of the models we had the following list:

**Table 4.8:** Information Criterions

| Model | AIC | BIC |
|---|---|---|
| **OLS** | 1146.96 | 1158.06 |
| **SLX** | 1146.02 | 1164.52 |
| **SAR** | 1145.04 | 1157.99 |
| **SEM** | **1142.91** | **1155.86** |
| **SDM** | 1145.07 | 1165.42 |
| **SDEM** | 1145.60 | 1165.95 |
| **SARAR** | 1143.73 | 1158.53 |
| **MANSKI** | 1147.06 | 1169.26 |

The model with the smallest AIC/BIC value was the Spatial Error Model (SEM) and thus proofed the fit was adequate.

### 4.4.3 Spatial Hausman Test

The test was conducted to find out if the OLS model was affected by endogeneity.

**Table 4.9:** Spatial Hausman Test

| Hausman test (asymptotic) | Degrees of freedom | p-value |
|---|---|---|
| 3.38 | 5 | **0.64** |

The test was not significant and thus there were no endogenous regressors in the model that could have caused the OLS to fail. Therefore, the outcome of the SEM was as a result of the random effects present.

### 4.5 Thematic Maps

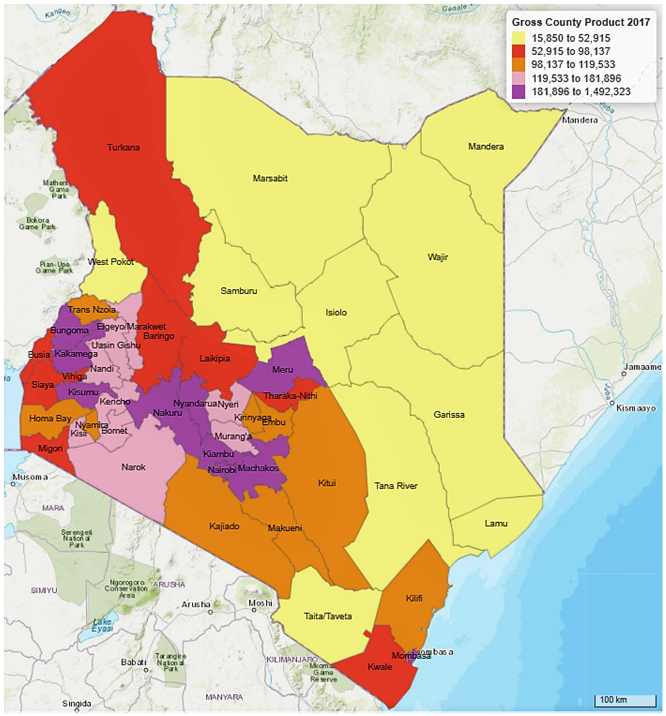



**Figure 4.9:** Total Gross County Product Map

The **Gross county product** (TOTAL) in millions of Kenyan shillings was widely spread through the 47 counties with large values concentrated in the Nairobi, Kiambu, Nakuru and Mombasa Counties revealing the uneven economic distribution among the counties. The variation in colors according to categories showed the disparity that existed in each of the indicators among Kenyan counties.

**Constr** meant construction of Building plans approved and their value, Value of completed buildings, Fees from building permits/approvals. Large values were in Nairobi and Kiambu counties. The rest of the counties were evident from the map.

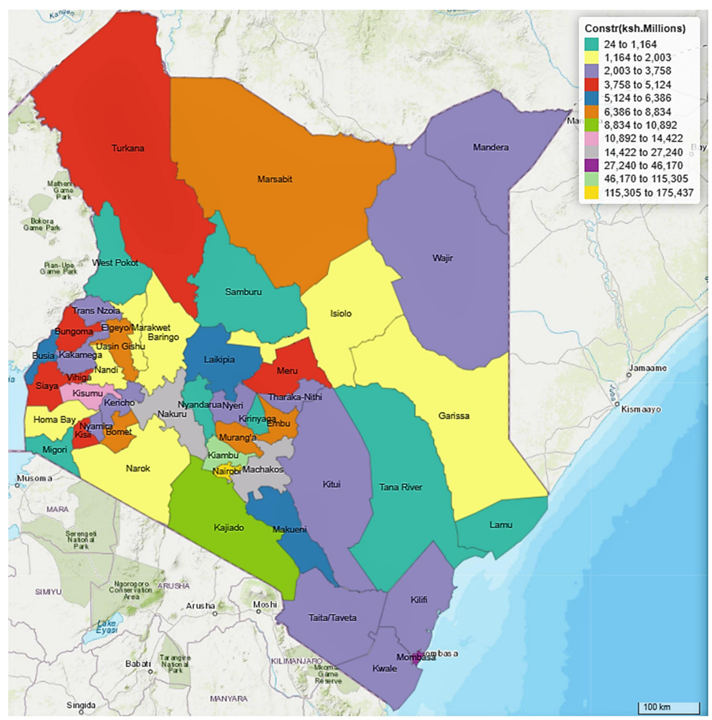



**Figure 4.10:** Map of Kenya showing the GCP indicator Construction.

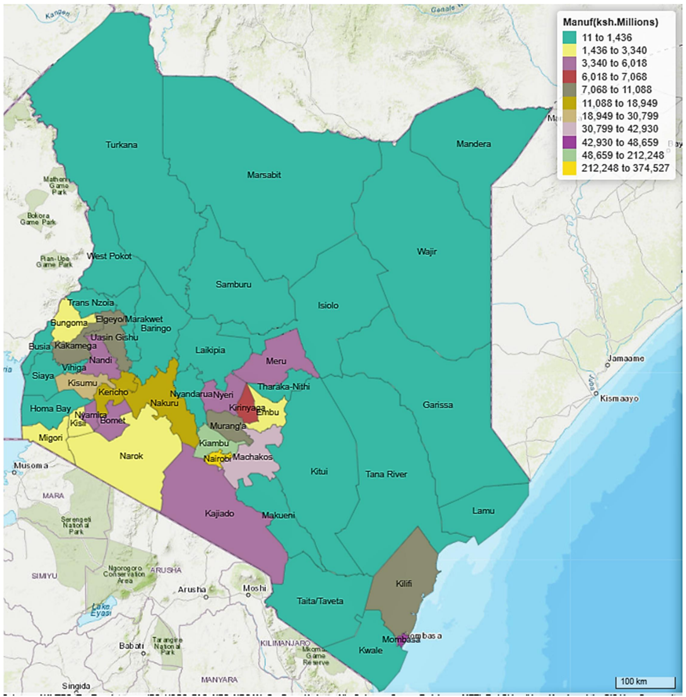



**Figure 4.11:** Map of Kenya showing the GCP indicator manufacturing.

**Manuf** meant Manufacturing of Food, beverages and tobacco, Non-food products and Repairs which was highly concentrated in Nairobi and Kiambu counties. The rest of the counties were evidently displayed in Figure 4.11.

**FinInsuA** meant Financial and insurance activities which included Insurance, reinsurance and pension funding, activities auxiliary to financial service, other financial activities. High values were observed in Nairobi and Kiambu County as from Figure 4.12.

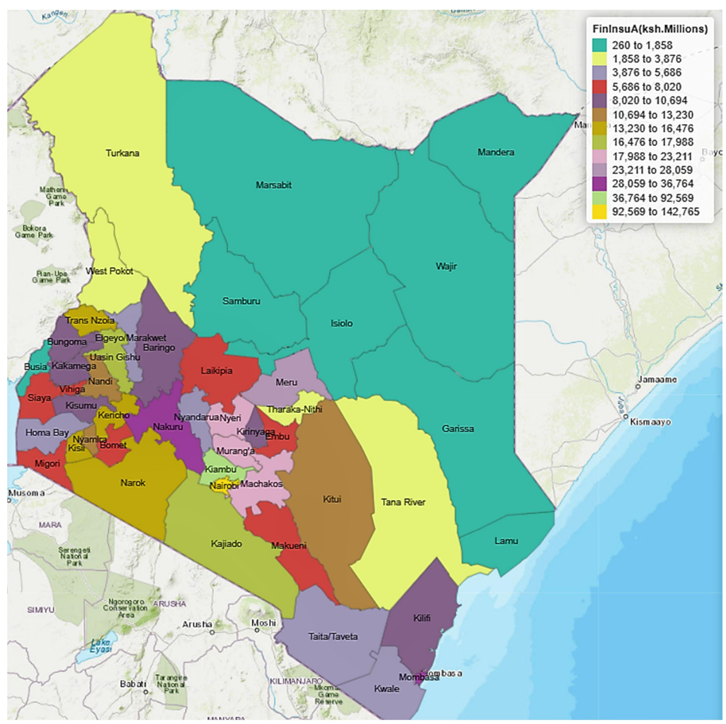



**Figure 4.12:** Map of Kenya showing the GCP indicator Financial and Insurance Activities

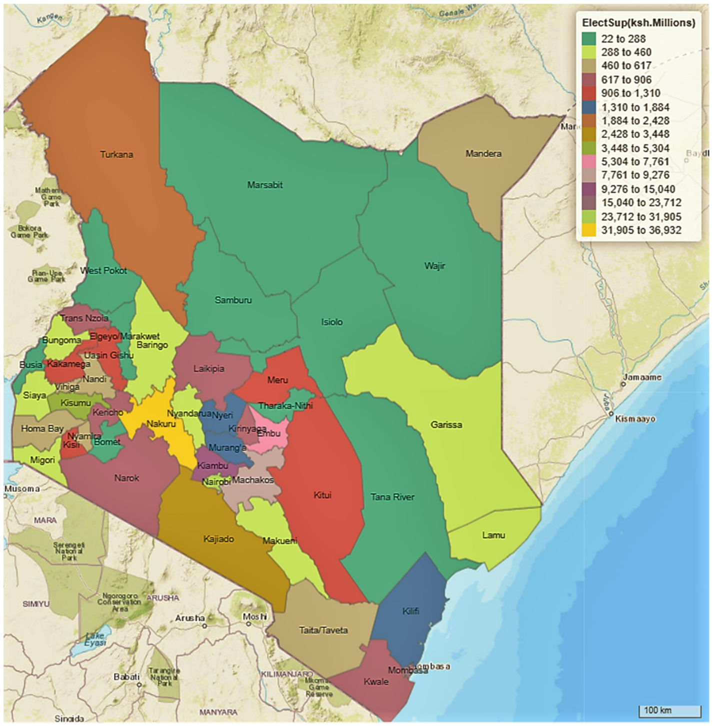



**Figure 4.13:** Map of Kenya showing the GCP indicator Electricity Supply

**ElectSup** in Figure 4.13 meant Electricity supply and included Power generation, Power transmission and Power distribution. Low values were observed in Tana River, Wajir, West Pokot and Samburu counties.

**AgrForFish** in Figure 4.14 meant Agriculture, forestry and fishing which included Growing of crops, Use of farm inputs, Animal production, Support services, Forestry and logging, Fishing and other fishing products with large values in Nakuru, Nyandarua, Kiambu and Elgeiyo Marakwet.

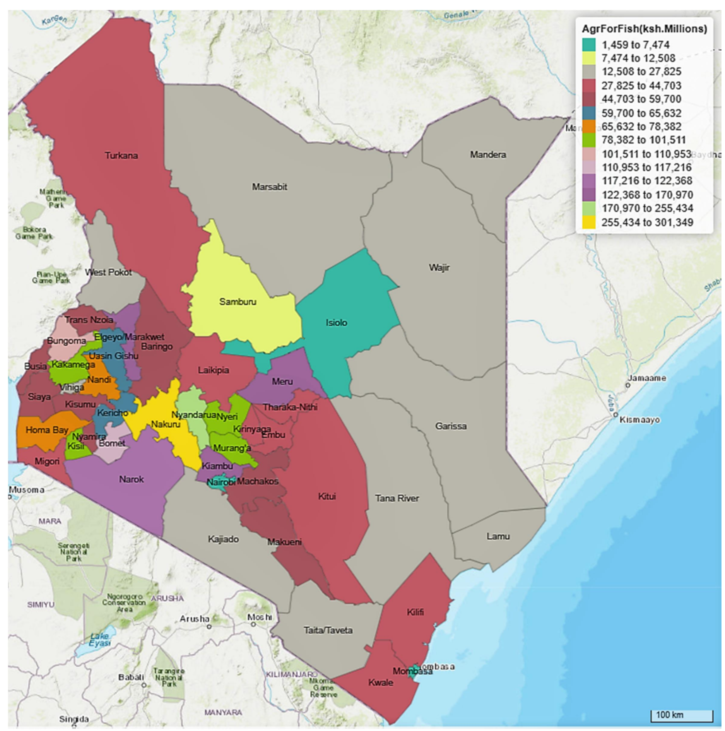



**Figure 4.14:** Map of Kenya showing the GCP indicator Agriculture, forestry and fishing

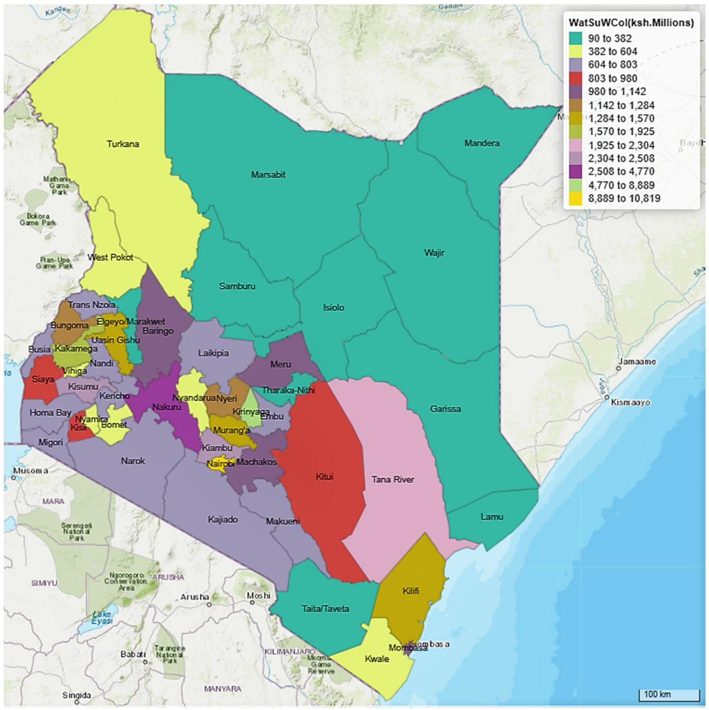



**Figure 4.15:** Map of Kenya showing the GCP indicator Water supply and waste collection

**WatSuWCol** in Figure 4.15 meant Water supply and waste collection. It included Water supply and Sewerage, Waste collection and treatment which were clearly based in large values at Nairobi and Kirinyaga counties.

**WholRRMV** in Figure 4.16 meant **Wholesale and retail trade; repair of motor vehicles** which included their Sales, Retail sales and Vending/hawking. Large values were observed in Nairobi and Mombasa counties.

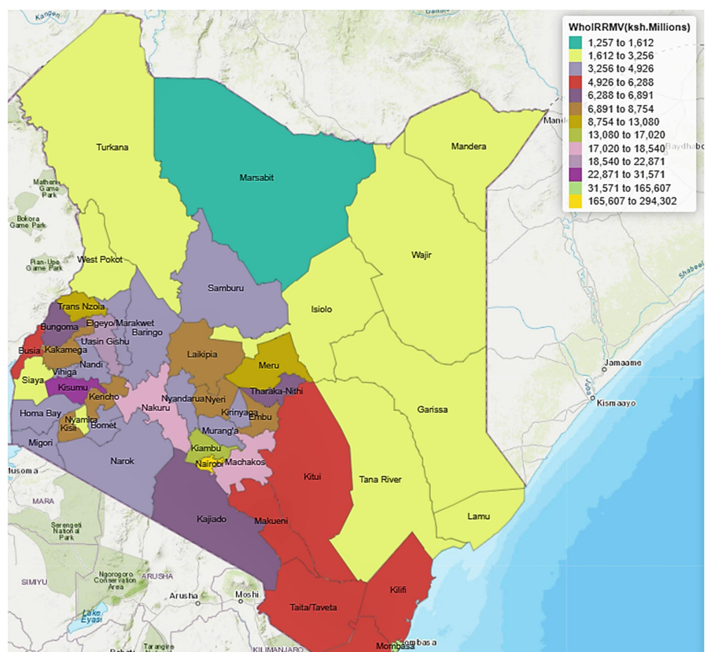



**Figure 4.16:** Map of Kenya showing the GCP indicator Wholesale and retail trade; repair of motor vehicles

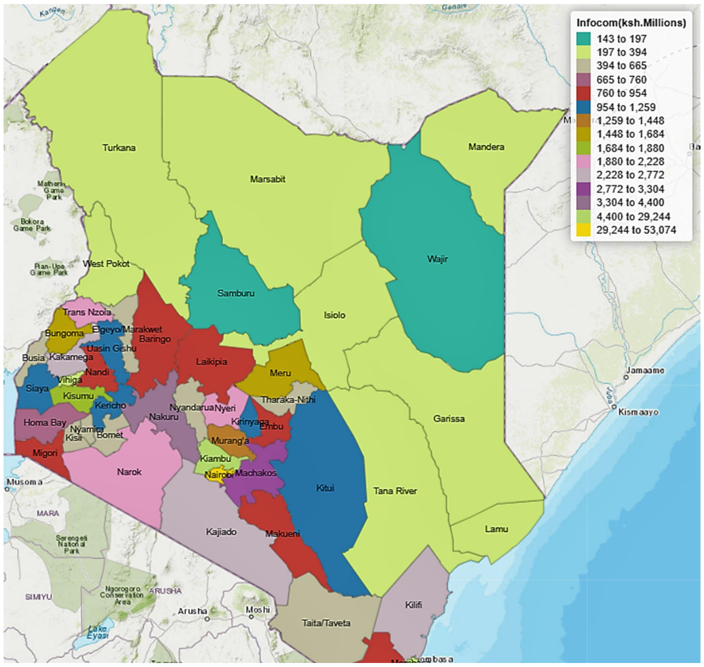



**Figure 4.17:** Map of Kenya showing the GCP indicator Information and Communication

**Infocom** in Figure 4.17 meant **Information and communication** which included Tele-communications, IT and other Information service activities. Large values were observed in Nairobi, Mombasa and Kiambu counties.

**RelEstA** in Figure 4.18 meant Real estate rental activities. Large values were observed at Nairobi, Mombasa and Kiambu counties.

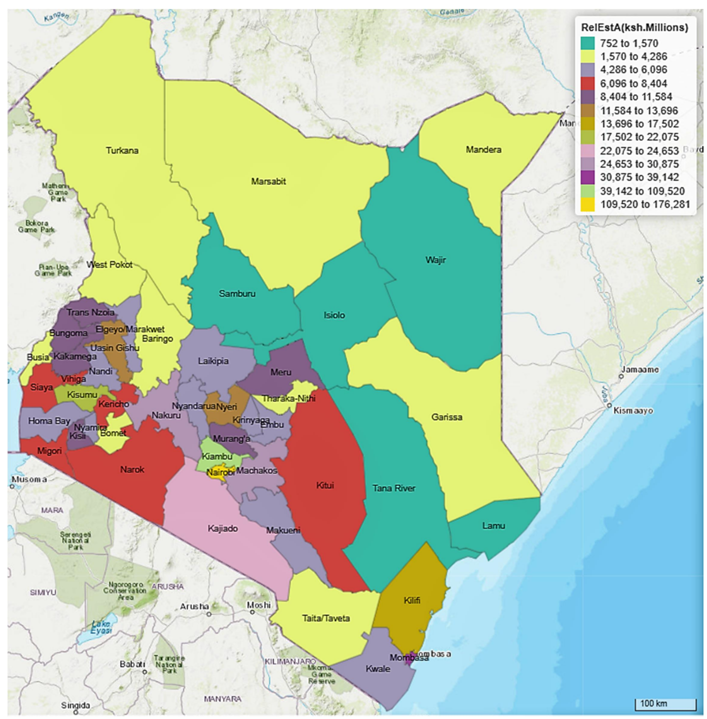



**Figure 4.18:** Map of Kenya showing the GCP indicator Real estate rental activities

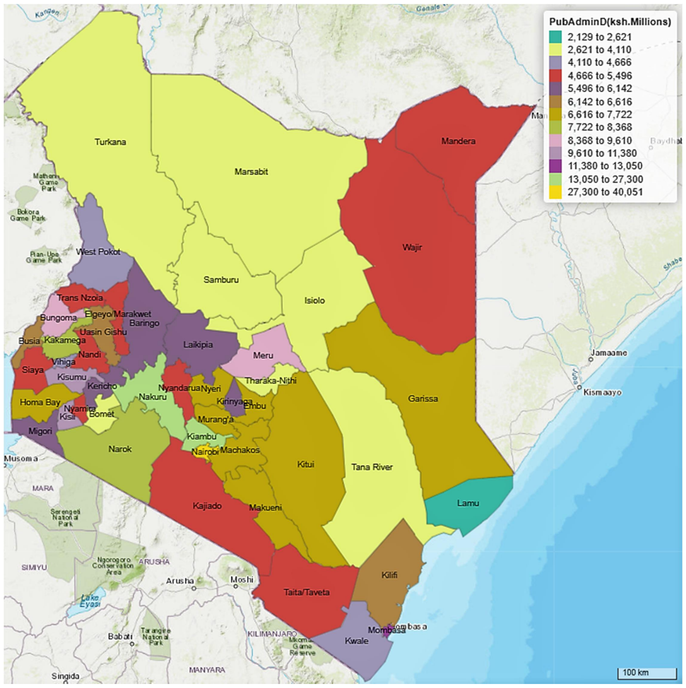



**Figure 4.19:** Map of Kenya showing the GCP indicator Public administration and Defence

**PubAdminD** in Figure 4.19 meant **Public administration and Defence** which included compulsory social security. Low values were observed in Marsabit, Lamu, Tana River and Tharaka Nithi.

**AccoFodSA** in Figure 4.20 meant **Accommodation and food service activities** which consisted of Hotels, Other accommodation facilities, Number of employees, restaurants, cafes, food kiosks and others. Low values were observed in Marsabit, Mandera, Wajir and Tana River counties.

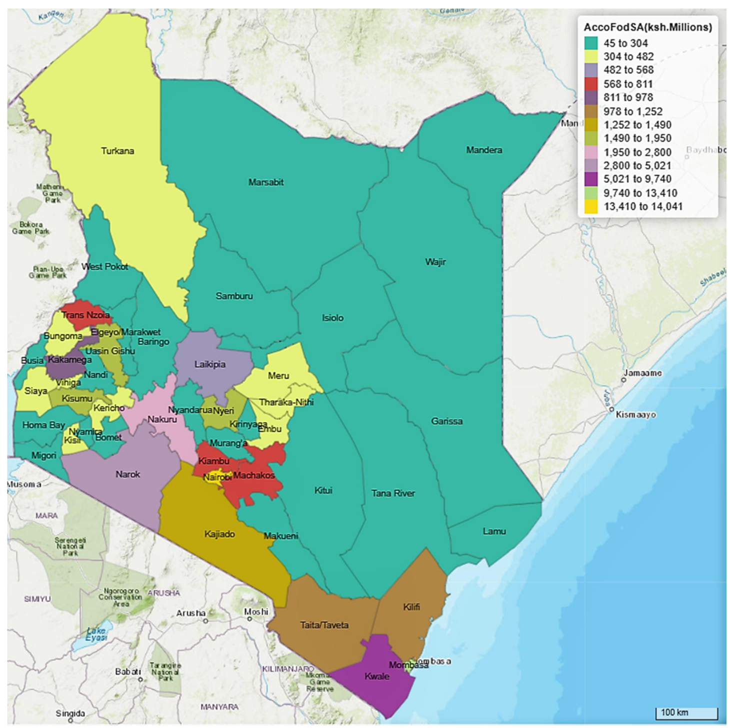



**Figure 4.20:** Map of Kenya showing the GCP indicator Accommodation and food service activities

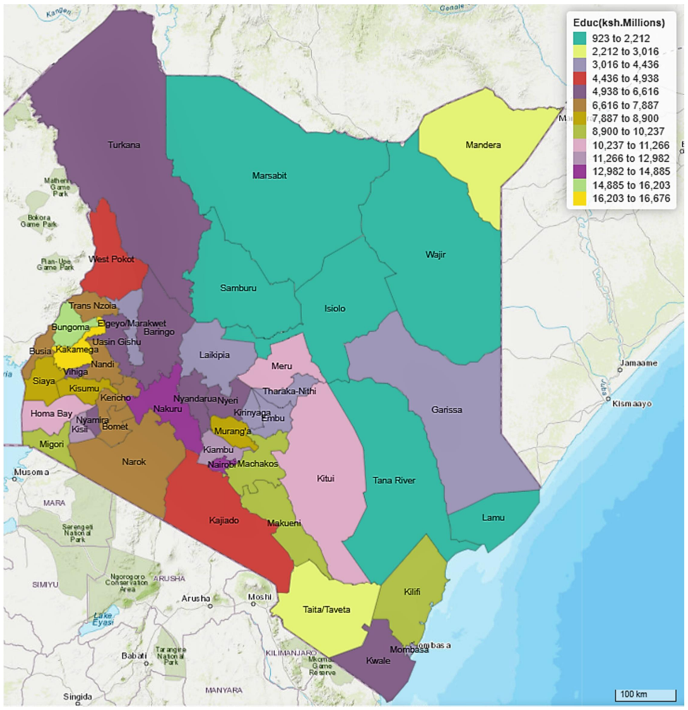



**Figure 4.21:** Map of Kenya showing the GCP indicator Education

**Educ** in Figure 4.21 meant **Education** which included Pre-primary, Primary, General Secondary, Technical Vocational Education and Training Institutions, Higher Education and Other education. Large values were found in Kakamega and Bungoma counties.

**TranSto** in Figure 4.22 meant **Transport and storage** which included Land transport, water transport, other related activities, warehousing and storage. Small values were observed in Marsabit, Wajir, Isiolo and Tana River counties.

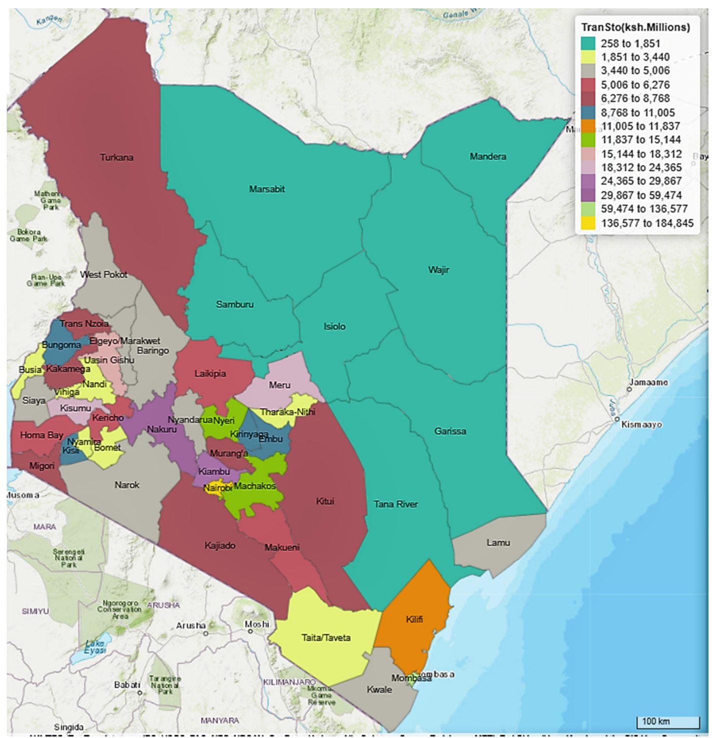



**Figure 4.22:** Map of Kenya showing the GCP indicator Transport and storage

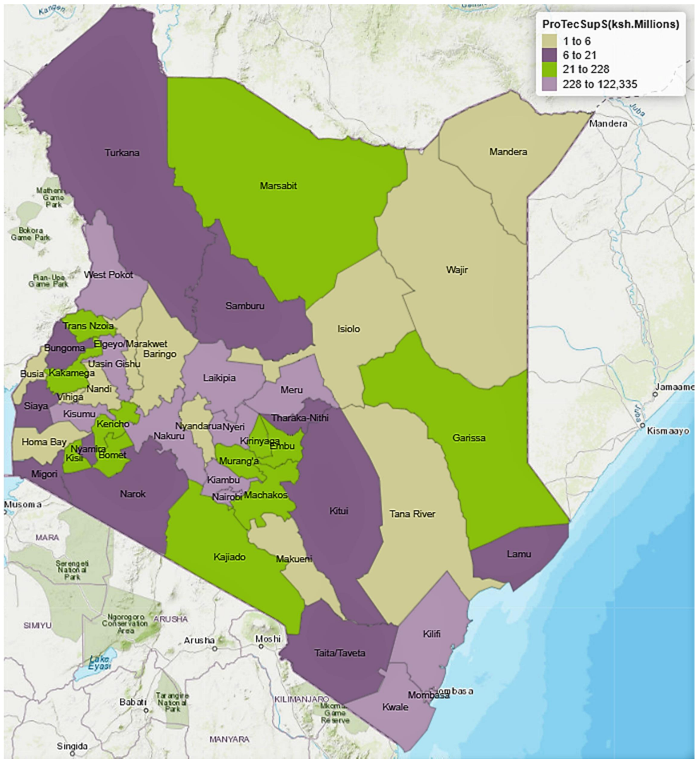



**Figure 4.23:** Map of Kenya showing the GCP indicator Professional, technical and support services

**ProTecSupS** in Figure 4.23 meant **Professional, technical and support services** which included Professional, scientific and technical activities. Large values observed in Nairobi and Mombasa Counties.

**MinQua** in Figure 4.24 meant Mining and quarrying which consisted of Quarrying, Sand Harvesting, Mineral exploitation, Gemstones, other minerals and Mineral production. Large values found in Machakos, Kiambu, Kilifi, Migori, Meru and West Pokot.

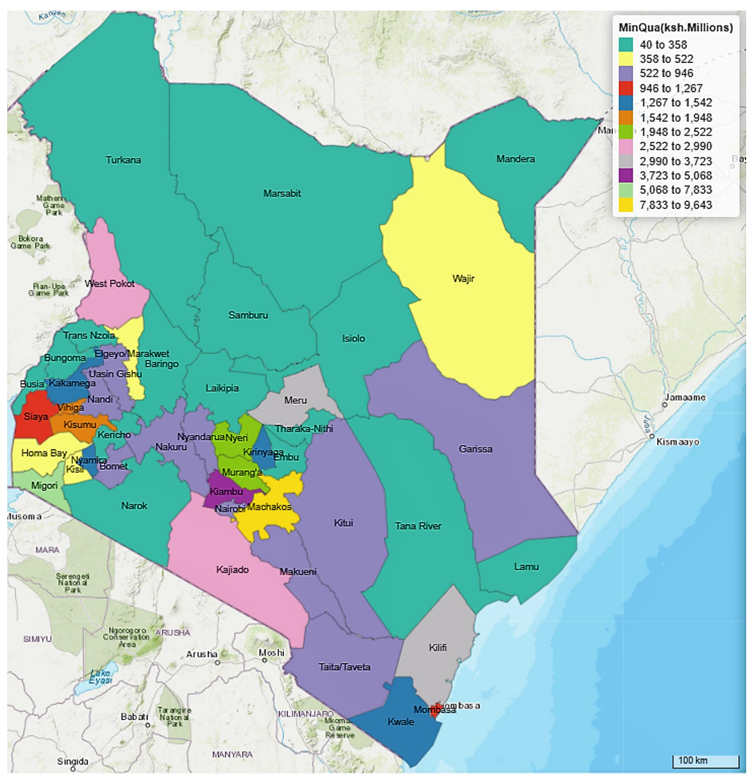



**Figure 4.24:** Map of Kenya showing the GCP indicator Mining and quarrying

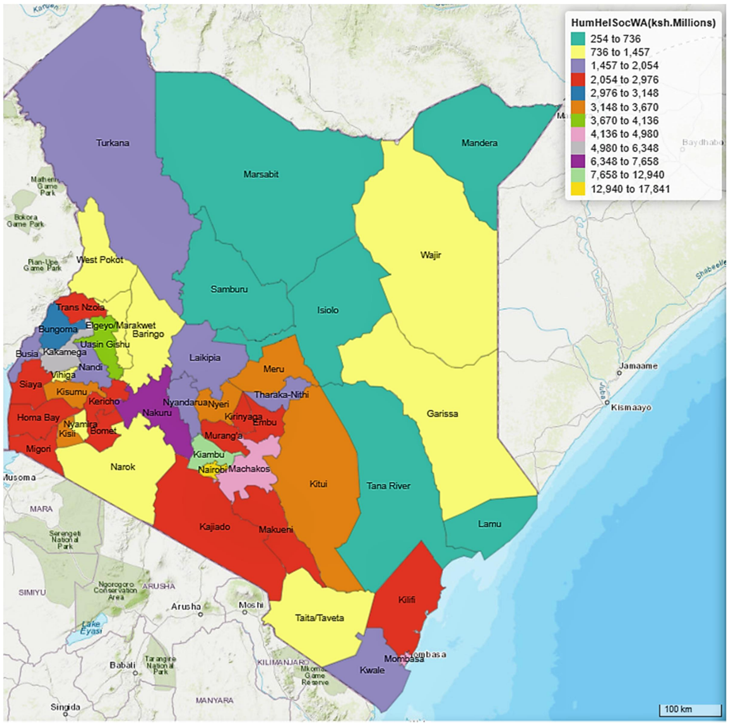



**Figure 4.25:** Map of Kenya showing the GCP indicator Human health and social work activities

**HumHelSocWA** in Figure 4.25 meant **Human health and social work activities** which included Hospitals, Health centers/Clinics/Dispensaries, staffing, arts, entertainment, recreation, services of membership organizations and other medical facilities. Low values observed in Samburu, Isiolo, Tana River and Lamu counties.

**Other services** in Figure 4.26 meant other related activities that are not part of the other variables like other entities had larger values in the Nairobi and Nakuru counties.

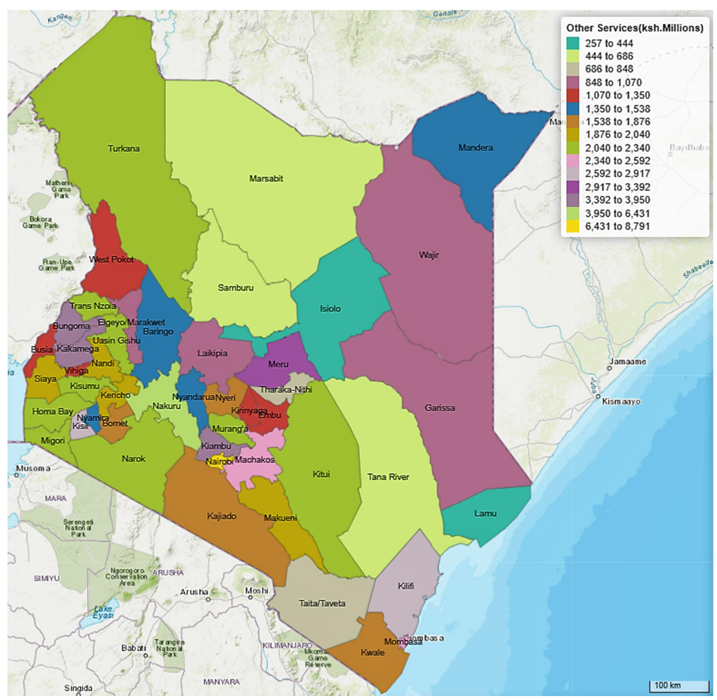



**Figure 4.26:** Map of Kenya showing the GCP indicator other service activities

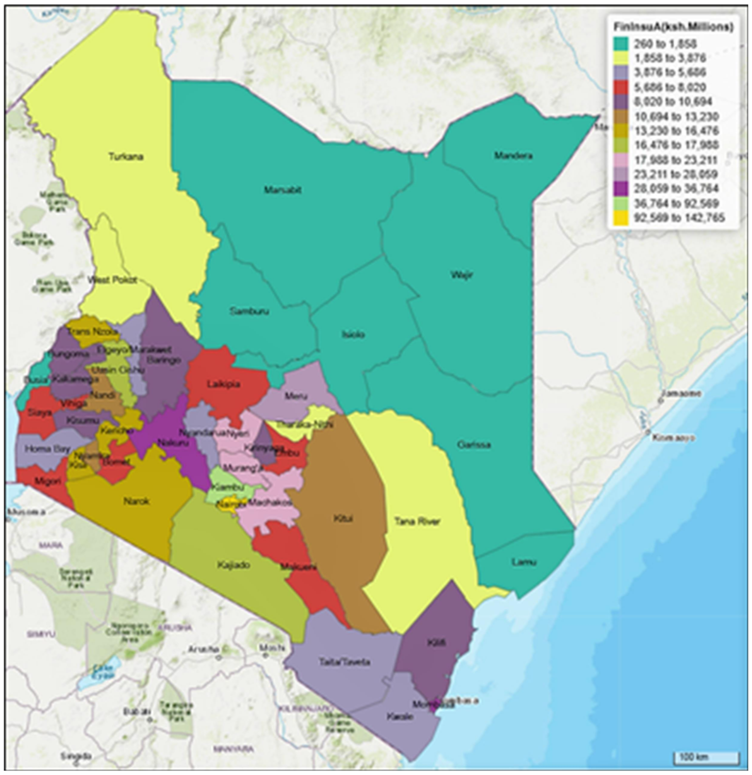



**Figure 4.27:** Map of Kenya showing the GCP indicator financial intermediation services indirectly measured

**FinIServM1** in Figure 4.27 meant **financial intermediation services indirectly measured** which included monetary intermediation which is subtracted. Large values were observed in Nairobi, Mombasa and Uasin Gishu counties.

| Economic Bloc (Number of Counties) | Counties |
|---|---|
| Frontier Counties Development Council (FCDC) (7) | Garissa, Wajir, Mandera, Isiolo, Marsabit, Tana River and Lamu |
| North Rift Economic block (NOREB) (8) | Uasin Gishu, Trans-Nzoia, Nandi, Elgeyo Marakwet, West Pokot, Baringo, Samburu and Turkana |
| Lake Region economic block (LREB)(14) | Migori, Nyamira, Siaya, Vihiga, Bomet, Bungoma, Busia, Homa Bay, Kakamega, Kisii, Kisumu, Nandi, Trans Nzoia and Kericho |
| Jumuia ya kaunti za pwani (6) | Tana River, Taita Taveta, Lamu, Kilifi, Kwale and Mombasa |
| South Eastern Kenya Economic Block (3) | Kitui, Machakos and Makueni. |
| Mt.Kenya and Aberdares Region Economic block AKA Central Kenya Economic Bloc (CEKEB) (10) | Nyeri, Nyandarua, Meru, Tharaka Nithi, Embu, Kirinyaga, Murang'a, Laikipia, Nakuru and Kiambu |

**Figure 4.28:** Kenyan economic blocs' summary

The Gross County Product was not distributed in a pattern as displayed by the economic blocs. The spatial shock observed revealed an association in space of the counties as shown in Figure 4.7 and suggested on how the economic blocs should have been arranged.

# CHAPTER FIVE

# SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Introduction to summary

This chapter presented the summary of the results, conclusions and recommendations from the results of the geospatial data analyzed. The objectives were all a priority and explanations were given on those that did not need to be performed extensively.

## 5.2 Summary of the research

The spatial dependence of the Gross county product (GCP) of Kenya on its indicators was clearly investigated as a new strategy of describing the economic performance of Kenyan counties impartially. The GCP was a regional economic measure which disaggregates the GDP.

The GCP was necessary for:

i. Estimation of revenue potential for each county.
ii. Informing economic progress at the county level.
iii. An indicator for potential for private sector investment.
iv. Informing county economic development plans.

The background of the study was based on spatial effects and how non-spatial models have been pioneering geo-economic research areas globally causing model misspecification. It targeted on the lack of geo-spatial modelling in Kenya and the last of such studies concerning Kenyan regions was in 2007 during provincial boundaries. Also by appreciating the efforts of the Kenya Accountable Devolution Program (KADP) who after producing GCP values in 2019, it discussed on how it could be wasteful if they had not yet engaged into spatial econometrics.

The literature review in the second chapter targeted several authors and countries that had approached regional spatial research in other fields apart from the economic fraternity. It highlighted in details the 7 spatial models selected for this study, how different authors had engaged them to observe temporal or spatial effects and how they had been excluded in the regional studies of Kenya. The research used a quantitative and cross-sectional approach of deriving spatial estimates and their statistical significances.

The study area consisted of the 47 counties of Kenya and the findings capturing the state of each economic bloc. The current price of goods and services in millions of Kenyan shillings as on 2019 from KNBS was used as the secondary dataset. The data was analyzed in R (version 4.3.0) and ArcGIS (version 10.7.1) and findings presented in the fourth chapter by making use of maps, statistical diagrams, figures and tables.

The findings revealed that the suitable model to define the relationship between the GCP and the 18 indicators was the Spatial Error Model. From the model, it was observed that there was a residual pattern represented as a vector of auto correlated disturbances $\mu = [(I_{47} - 0.52736W)^{-1}38399] \times 10^{6}$ which was causing the spatial phenomenon that was not only in a single place but among the regions and beyond. For example, fish moving from Lake Turkana in Turkana County in a truck so as to be sold in Nairobi County would face such an effect which would be progressive and accumulative even beyond Kenyan borders. This error model was the one that accounted for the spatial autocorrelation observed which can be attributed to the inflatory state of goods and services and also the net exports as counties receive and dispatch them between one another and abroad at a different price rather than the current one.

To reduce the number of dimensions, tests were done on the indicator dataset. The Kasier – Meyer –Olkin (KMO) validity test for the factor analysis produced a value of **0.89** which was meritorious. Bartlett's test of Sphericity produced a chi-square value of **545.56** which was significant with p-value of **1.7373** thus making the factor analysis acceptable. The reliability value of the data for factor analysis was the Cronbach coefficient α=**0.8862** which was above the standard value of 0.7.After conducting the factor analysis procedure, the 18 indicator variables were reduced to 4 factors through the maximum likelihood method and the factor models were compared using an ANOVA. The one with 4 factors was chosen because it had the least value of the BIC=**-159.32.** Factor scores were obtained from the 4-factor model and labelled as latent variables for the next objective. Then a Moran I test revealed the presence of spatial autocorrelation which was positive at **0.2118** and with the p-value = **0.0024.**The Lagrange Multiplier/ Rao score test also gave LMerr as statistically significant with a p-value **0.03044** * at an alpha level of 0.05.The Spatial Error model therefore, had a p-value of **0.0139**, a lambda value of **0.5273** and an error term $\varepsilon$ of **38399,** BIC value of **1155.8570** and AIC value of **1142.906** which were the least among the rest of the spatial models**.** The model was given as;

TOTAL = 159809.2+192073.2 ML1+30559.6 ML2+38472.3 ML3-22577.3ML4+$(I_{47} - 0.53W)^{-1}38399$

A Likelihood Ratio test showed that the SEM was not to be restricted to simpler forms according to Table 4.8 and the Spatial Hausman test gave a value of **3.38** (p-value=**0.64)** which was not significant thus no endogenous regressors.

There was no marginal effect for the Spatial Error Model. Though a Spatial lagged X model was done, the lags produced estimates that were not significant thus their marginal effect was suppressed. Each of the 18 indicators was mapped through R software tmap package and the outcome was presented in previous chapter.

### 5.3 Conclusion

The analysis of the GCP variable using spatial regression models had brought about new opinions as shown in the summary about the economic performance of the regions of Kenya. This study should be a benchmark for monitoring economic performance and direction. The annual domestic product should be discussed broadly based on the concepts outlaid by the study. For the 2019 GCP geocoded dataset, the suitable model to explain the relationship between the indicators and the GCP values was the Spatial Error Model. It was easy for stakeholders of the GCP spatial analysis to obtain map details from the R commands given in the appendix B table.

### 5.4 Recommendations

Further research was recommended on the; avoid bulleting in the thesis

i. Use of other proxies to define the regional spatial economic model in Kenya.
ii. Use of complex spatial models, spatial temporal models and other non-parametric spatial models for future geocoded datasets.
iii. Continuous production of regional economic data annually to monitor and produce spatial econometric reports.
iv. Utilization of spatial autocorrelation results to illustrate the current Kenyan economic blocs as displayed in Figure 4.7.
v. Creation of equal economic opportunities among Kenyan counties by stretching the benefits of the effects of the Gross County Product from Nairobi to all other 46 regions to avoid economic biases.

vi. Need for county governments to seek the wellbeing of their economy by doing detailed economic research.

## REFERENCES


Afifah, D. L., Rahardjo, S., Kusumasari, V., Atikah, N., & Kholifia, N. (2021). Optimization of tax revenue in East Java: An empirical analysis of spatial Durbin error model. *Journal of Physics: Conference Series*, *1872*(1), 012025. https://doi.org/10.1088/1742-6596/1872/1/012025

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723. https://doi.org/10.1109/TAC.1974.1100705.

Akhter, Y., Mahsin, M. D., & Mohaimin, M. Z. (2012). An application of factor analysis on gross domestic product data of Bangladesh. *Bangladesh e-Journal of Sociology*, *9*(1), 6-18.

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science*, *92*(1), 3-17. https://doi.org/10.1111/j.1435-5957.2012.00480.x

Anselin, L., & Florax, R. J. G. M. (1995). Small sample properties of tests for spatial dependence in regression models: some further results. In L. Anselin, & R. J. G. M. Florax (Eds.), *New directions in spatial econometrics* (pp. 21-74). Springer. https://doi.org/10.1007/978-3-642-79877-12

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, *20*(1), 1-17. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of The Academy of Marketing Science*, *16*,74-94. https://doi.org/10.1177/009207038801600107

Bai, A., Hira, S., & Deshpande, P. S. (2015). An application of factor analysis in the evaluation of country economic rank. *Procedia Computer Science*, *54*, 311-317. https://doi.org/10.1016/j.procs.2015.06.036

Bai, C. E., Ma, H., & Pan, W. (2012). Spatial spillover and regional economic growth in China. *ChinaEconomicReview*, *23*(4),982990. https://doi.org/10.1016/j.chieco.2012.04.016

Balash, V., Balash, O., Faizliev, A., & Chistopolskaya, E. (2020). Economic growth patterns: Spatial econometric analysis for Russian regions. *Information, 11*(6), 289. http://dx.doi.org/10.3390/info11060289

Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatialdata analysis with R series*. (*Vol. 747248717*, pp. 237-268). Springer. http://dx.doi.org/10.1007/978-0-387-78171-6

Burkey, M. L. (2018). Spatial econometrics and GIS YouTube playlist. *Region 5*(3), R13-R18. https://doi.org/10.18335/region.v5i3.254

Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A*. *13*(7), 795-800. https://doi.org/10.1068/a130795

Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science, 52*(2), 210-239.https://doi.org/10.1111/j.1467-9787.2011.00726.x

Cox, T. F. (1984). [Review of Spatial Processes*: Models and Applications,* by A. D. Cliff & J. K. Ord]. *Journal of the Royal Statistical Society. Series A (General)*, *147*(3), 515–515. https://doi.org/10.2307/2981591

Cronbach, L.J.(1951).Coefficient alpha and the internal structure of tests. P*sychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis*. Sage. https://doi.org/10.1111/j.1538-4632.2001.tb00453.x

George, D., & Mallery, P. (2021). IBM SPSS statistics 27 step by step. https://doi.org/10.4324/9781003205333

Huang, Y. (2014). Drivers of rising global energy demand: The importance of spatial lag and error dependence. *Energy*, *76*, 254-263. http://dx.doi.org/10.1016/j.energy.2014.07.093

Jiang, L., Folmer, H., Ji, M., & Tang, J. (2017). Energy efficiency in the Chinese provinces: A fixed effects stochastic frontier spatial Durbin error panel analysis. *The Annals of Regional Science*, *58*(2), 301-319. https://doi.org/10.1007/s00168-016-0782-5

Karim, A., Faturohman, A., Suhartono, S., Prastyo, D. D., & Manfaat, B. (2017). Regression models for spatial data: An example from Gross Domestic Regional Bruto in Central Java. *Jurnal ekonomi pembangunan: Kajian masalah ekonomi dan pembangunan, 18*(2), 213-224. https://doi.org/10.23917/jep.v18i2.4660

Kanaroglou, P. S., Adams, M. D., De Luca, P. F., Corr, D., & Sohel, N. (2013). Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressivemodel. *Atmospheric Environment*, *79*,421-427. https://doi.org/10.1016/j.atmosenv.2013.07.014

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, *157*(1), 53-67. https://doi.org/10.1016/j.jeconom.2009.10.025

Kenya National Bureau of Statistics (KNBS) (2019). *Gross County Product Report 2019* [PDF] (No.978-9966-102-08-9) https://africaopendata.org/dataset/kenya-economic-survey-2019.

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference*. Cambridge university press. https://doi.org/10.3998/mpub.23784

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management*, *45*(1), 24-39. https://doi.org/10.1016/S0095-0696(02)00013-X

Kwame, O. (2019). *Four lessons from the publication of the county gross product* [Daily Nation Blog]. https://nation.africa/kenya/blogs-pinion/blogs/dot9/kwame/four-lessons-from-the publication-of-the-county-gross-product-138932.

LeSage, J. P., & Pace, R. K. (2009). Spatial econometric models. *Handbook of Applied Spatial Analysis*, 355-376. https://doi.org/10.1007/978-3-642-03647-7_18

LeSage, J.P. (2015). Theory and practice of spatial econometrics. *Spatial Economic Analysis, 10*, 400-400. https://doi.org/10.1080/17421772.2015.1062285

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531-542. https://doi.org/10.2307/2298123

McMillen, D. P. (2012). Perspectives on spatial econometrics: linear smoothing with structured models. *Journal of Regional Science*, *52*(2), 192-209. https://doi.org/10.1111/j.1467-9787.2011.00746.x

Mitra, S. K., & Saphores, J. D. M. (2016). The value of transportation accessibility in a least developed country city – The case of Rajshahi City, Bangladesh. *Transportation Research Part A: Policy and Practice*, *89*, 184-200. https://doi.org/10.1016/j.tra.2016.05.002

Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23. https://doi.org/10.2307/2332142

Moreno, R.,& Trehan, B. (1997). Location and the growth of nations. *Journal of Economic Growth*, *2*(4), 399-418. https://doi.org/10.1023/A:1009741426524

Ngugi, B. (2019). *Nairobi, Nakuru and Kiambu Counties Top Contributors to Kenya's Wealth.* biasharaleo.co.ke. https://biasharaleo.co.ke/nairobi-nakuru-and-kiambu-counties-top-contributors-to-kenyas-wealth/

Nunnally, J. C.,& Bernstein, IH (1994). The assessment of reliability. *Psychometric theory. 3* (1), 248-292. https://doi.org/10.12691/education-5-5-2

Okwi, P. O., Ndeng'e, G., Kristjanson, P., Arunga, M., Notenbaert, A., Omolo, A. & Owuor, J. (2007). Spatial determinants of poverty in rural Kenya. *Proceedings of the National Academy of Sciences*, *104*(43),16769-16774. https://doi.org/10.1073/pnas.0611107104

Olubusoye, O. E., Korter, G. O., & Salisu, A. A. (2016). Modeling road traffic crashes using spatial autoregressive model with additional endogenous variable. *Statistics in Transition New Series*, *17*(4), 659-670. https://doi.org/10.21307/stattrans-2016-045

Oud, J. H., & Folmer, H. (2008). A structural equation approach to models with spatial dependence. *Geographical Analysis*, *40*(2), 152-166. https://doi.org/10.1111/j.1538-4632.2008.00717.x

Paelinck, J. H. P. (1978). Spatial econometrics. *Economics Letters*, 1(1), 59-63. https://doi.org/10.1016/0165-1765(78)90097-6

Páez, A., Le Gallo, J., Buliung, R. N., & Dall'Erba, S. (2009). Progress in spatial analysis: Introduction. *Advances in Spatial Science*, 1-13. https://doi.org/10.1007/978-3-642-03326-1_1

Pandit, R., Polyakov, M., & Sadler, R. (2014). Valuing public and private urban tree canopy cover. *Australian Journal of Agricultural and Resource Economics*, *58*(3), 453-470. https://doi.org/10.1111/1467-8489.12037

Polyakov, M., Pannell, D. J., Pandit, R., Tapsuwan, S., & Park, G. (2015). Capitalized amenity value of native vegetation in a multifunctional rural landscape. *American Journal of Agricultural Economics*, *97*(1), 299-314. https://doi.org/10.1093/ajae/aau053

Pozdílková, A., & Marek, J. (2018). Spatial lag model for apartment prices in Pardubice region. In *17th Conference on Applied Mathematics APLIMAT 2018: proceedings*. Slovenská technická univezita v Bratislave. https://hdl.handle.net/10195/72924

Ramirez, M. T., & Loboguerrero, A. M. (2002). Spatial dependence and economic growth: Evidence from a panel of countries. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.311320

Revelle, W. (2019). *PsychTools: Tools to accompany the 'psych' package for psychological research.* CRAN:ContributedPackages. https://doi.org/10.32614/cran.package.psychtools

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Septiawan, A., Handajani, S., & Martini, T. (2018). Spatial Durbin error model for human development index in province of Central Java. *Journal of Physics: Conference Series*, *1025*, 012107. https://doi.org/10.1088/1742-6596/1025/1/012107

Syed, A. A. S. G., & Shaikh, F. M. (2013). Effects of macroeconomic variables on gross domestic product (GDP) in Pakistan. *Procedia Economics and Finance*, *5*, 703-711. https://doi.org/10.1016/S2212-5671(13)00082-8

Szendi, D. (2016). Spatial beta convergence analysis of the real GDP per capita across Germany and Hungary. *Doktornduszok Fóruma Szekciókiadván.* ResearchGate. https://orcid.org/0000-0003-0010-9949

Tientao, A., Legros, D., & Pichery, M. C. (2016). Technology spillover and TFP growth: A spatial Durbin model. *International Economics*, *145*, 21-31. https://doi.org/10.1016/j.inteco.2015.04.004

Xiangyu, G., Zhou, Z., Zhou, Y., Ye, X., & Liu, S. (2018). A spatial panel data analysis of economic growth, urbanization, and NOx emissions in China. *International Journal of Environmental Research and Public Health*, *15*(4), 725. https://doi.org/10.3390/ijerph15040725

Xiao, H., & Mao, J. (2020). Effects of postgraduate education on technological innovation: A study based on the spatial Durbin model. *Asia Pacific Education Review*, *22*(1), 89-99. https://doi.org/10.1007/s12564-020-09652-y

Yoo, D., & Steckel, R. (2010). *Property rights and financial development: The legacy of Japanese colonial institutions*. https://doi.org/10.3386/w16551

Zivot, E. (2009, November 15). *Maximum Likelihood Estimation* [PDF]. University of Washington. http://faculty.washington.edu/ezivot/econ583/mleLectures.pdf

## APPENDICES

### APPENDIX A: Table of GCP Indicator Variables

| No. | Variables (Indicators) | Definition [Indicators from Expenditures, Production or Income as a measure of counties economy] | Type of Variable |
|---|---|---|---|
| 1. | **Educ** | **Education**: Pre-primary, Primary, General Secondary, Technical Vocational Education and Training Institutions, Higher Education and Other education. | Explanatory |
| 2. | **AgrForFish** | **Agriculture, forestry and fishing**: Growing of crops, Use of farm inputs, Animal production, Support services, Forestry and logging, Fishing and other fishing products. | Explanatory |
| 3. | **MinQua** | **Mining and quarrying**: Quarrying, Sand Harvesting, Mineral exploitation, Gemstones, other minerals and Mineral production. | Explanatory |
| 4. | **Manuf** | **Manufacturing**: Food, beverages and tobacco, Non-food products and Repairs. | Explanatory |
| 5. | **WatSuWCol** | **Water supply; waste collection**: Water supply and Sewerage, Waste collection and treatment. | Explanatory |
| 6. | **Constr** | **Construction**: Building plans approved and their value, Value of completed buildings, Fees from building permits/approvals. | Explanatory |
| 7. | **WholRRMV** | **Wholesale and retail trade; repair of motor vehicles**: Sales, Retail sales and Vending/hawking | Explanatory |
| 8. | **TranSto** | **Transport and storage**: Land transport, water transport, Other related activities, warehousing and storage. | Explanatory |
| 9. | **AccoFodSA** | **Accommodation and food service activities**: Hotels, Other accommodation facilities, Number of | Explanatory |

| | | | |
|---|---|---|---|
| | | employees, restaurants, cafes, food kiosks and others. | |
| 10. | **Infocom** | **Information and communication**: Tele-communications, IT and other Information service activities. | Explanatory |
| 11. | **FinInsuA** | **Financial and insurance activities**: Insurance, reinsurance and pension funding, activities auxiliary to financial service, other financial activities. | Explanatory |
| 12. | **RelEstA** | **Real estate activities**: Real estate rental activities. | Explanatory |
| 13. | **ProTecSupS** | **Professional, technical and support services**: Professional, scientific and technical activities. | Explanatory |
| 14. | **PubAdminD** | **Public administration and Defence**: Including Compulsory social security | Explanatory |
| 15. | **ElectSup** | **Electricity supply**: Power generation, Power transmission and Power distribution. | Explanatory |
| 16. | **HumHelSocWA** | **Human health and social work activities**: Hospitals, Health centers/Clinics/Dispensaries, staffing, arts, entertainment, recreation, services of membership organizations and Other medical facilities. | Explanatory |
| 17. | **Other Services** | **Other service activities**: Other related activities that are not part of the other variables like other entities. | Explanatory |
| 18. | **FinIServM1** | **Financial intermediation services indirectly measured**: Monetary intermediation which is subtracted. | Explanatory |
| 19. | **TOTAL** | **Total GCP**: The total Gross county product. | Response |

## APPENDIX B: Table of R Programming Commands

R Commands

```
fpacks=c("rgdal","rgeos","spdep","sf","dplyr","leaflet","tmap","tmaptools","ggplot2","
spatialreg","foreign","raster","RColorBrewer","maptools","DescTools","tidyverse","IS
LR","leaps","car","gridExtra","grid","fBasics","stats","olsrr",”corrplot”)
lapply(fpacks, require, character.only=TRUE)
GCP.data=readOGR(dsn=".", layer = "GCP20230" )
M<- cor(Eda[,10:27])
corrplot(M,method = "number")
sf_use_s2(FALSE)
tmap_mode("view")
GCP.data$ElectSup<- as.numeric(GCP.data$ElectSup)
map<- tm_shape(GCP.data)+tm_fill("ElectSup", palette = "Greens",style = "quantile",
title = "ElectSup(ksh.Millions)", n=15, legend.hist = TRUE) +tm_borders(alpha=.4)+
  tm_layout(title = "Electricity Supply \n in Kenya (Ksh.Millions) at Current Prices
2017 ") + tm_text("Name")+ tm_scale_bar(breaks = c(0, 2, 4), text.size = 1, position =
c("right", "bottom"))
print(map)
grid.newpage()
queen.nb=poly2nb(GCP.data)
queen.listw=nb2listw(queen.nb,style="W")##Use the listw stead of the poly2nb
listw1=queen.listw##You can change to rooks.
mat.list<- spdep::nb2mat(queen.nb, style = "W")
summary(listw1)#Is my matrix almost singular..or perfect correlation experienced
eigenw(listw1)
1/range(eigenw(listw1))
W <- as(listw1, "CsparseMatrix")
W
str(W)
all(W == t(W))# Asymmetric sparse matrices
n <- nrow(W)
set.seed(1)
x <- runif(n)
x
WX <- as.numeric(W %*% x)
plot(x, WX, ylim=c(0,1), las=1, pch=8, xlab="runif(n) values", main="The Sparse
```

```
Matrix W")
c(x[1], WX[1])
reg.eq1=TOTAL~ ML1+ML2+ML3+ML4
reg1<- lm(TOTAL~ ML1+ML2+ML3+ML4, data=GCP.data)
reg1
summary(reg1)
par(mfrow=c(2,2))
plot(reg1)
lm.morantest(reg1,listw1)##positive means clustering negative means dispersion##alternate hypothesis states that the data is spatially clustered
par(mfrow=c(1,1))
# creates a moran plot
moran <- moran.plot(reg1$residuals, listw = listw1, main="Moran Plot", xlab = "OLS residuals", ylab="Spatially Lagged residuals")
# creates a local moran output
local <- localmoran(x = GCP.data$TOTAL, listw = listw1)
local
# binds results to our polygon shapefile
moran.map <- cbind(GCP.data, local)
# maps the results
tmap_mode("view")
tm_shape(moran.map) +tm_text("Name")+ tm_fill(col = "Ii", midpoint = NA , style = "quantile", title = "local moran statistic")+tm_scale_bar(breaks = c(0, 2, 4), text.size = 1, position = c("right", "bottom"))
### to create LISA cluster map ###
quadrant <- vector(mode="numeric",length=nrow(local))
# centers the variable of interest around its mean
m.tot <- GCP.data$TOTAL - mean(GCP.data$TOTAL)
# centers the local Moran's around the mean
m.local <- local[,1] - mean(local[,1])
var(local[,1])
# significance threshold
signif <- 0.5
# builds a data quadrant
quadrant[m.tot >0 and m.local>0] <- 4
quadrant[m.tot <0 and m.local<0] <- 1
```

```
quadrant[m.tot <0 and m.local>0] <- 2
quadrant[m.tot >0 and m.local<0] <- 3
quadrant[local[,5]>signif] <- 0
# plot in r
brks <- c(0,1,2,3,4)
colors <- c("white","blue",rgb(0,0,1,alpha=0.4),rgb(1,0,0,alpha=0.4),"red")
plot(GCP.data,border="lightgray",col=colors[findInterval(quadrant,brks,all.inside=FA
LSE)])
box()
legend("bottomleft",legend=c("insignificant","low-low","low-high","high-low","high-
high"), fill=colors,bty="n")
p.values <- c(0.0001, 0.001, 0.006, 0.03, 0.095, 0.117, 0.234, 0.552, 0.751, 0.985)
tests<- lm.LMtests(reg1,nb2listw(queen.nb),test="all", zero.policy=TRUE)##the rlmlag
is significant man
p.adjust(tests$LMlag$p.value, method=p.adjust.methods)
#confint(tests, level = 0.1)
summary(tests)
AIC(reg1)
BIC(reg1)
#SLX spatially lagged model y=xB+WxT+e
reg2=lmSLX(reg.eq1, data=GCP.data, listw1)##SLX
summary(reg2)
impacts(reg2,listw=listw1)##The overall effect
summary(impacts(reg2,listw=listw1,R=500), zstats=TRUE)
AIC(reg2)
BIC(reg2)
betas<- coef(lm(formula =reg.eq1, data = GCP.data ))
pars= list(rho=0.5, sigma2e=2.0, betas=betas)
#res_res<- sar(reg.eq1, data=GCP.data, W=W,burnin = 500, Nsim = 1000, thinning =
1,parameters.start = pars)
#summary(res_res)
#Since the lagged y,s are global the error below could have ensued
reg3=lagsarlm(reg.eq1,data=GCP.data,listw1, tol.solve=1.0e-21, trs=NULL)##SAR or
The Lag y
summary(reg3)##In what way does p(rho) affect y and is it statistically significant
reg3b=spBreg_lag(reg.eq1,data=GCP.data,listw1,type="lag")
```

```
summary(reg3b)
impacts(reg3,listw=listw1)##The overall effect
summary(impacts(reg3,listw=listw1,R=500), zstats=TRUE)
AIC(reg3)
BIC(reg3)
##SEM model
reg4=errorsarlm(reg.eq1, data=GCP.data,listw1, tol.solve=1.0e-19)##SEM
summary(reg4)
AIC(reg4)
BIC(reg4)
Hausman.test(reg4)#Comparing SEM and OLS...SEM is the right model
reg5=errorsarlm(reg.eq1, data=GCP.data, tol.solve=1.0e-19,listw1, etype =
"emixed")##SDEM-Spatial Durbin Error Model
reg6=lagsarlm(reg.eq1, data = GCP.data,tol.solve=1.0e-40, listw1,
type="mixed")##SDM-Spatial Durbin Model
#reg7=sacsarlm(reg.eq1, data = GCP.data,tol.solve=1.0e-10, listw1,
type="sacmixed",method = "MC")##Manski All-inclusive Model--MINRES Method
reg7=sacsarlm(reg.eq1, data = GCP.data,tol.solve=1.0e-20, listw1,
type="sacmixed",method = "LU")##Manski All-inclusive Model--For MLE method
reg8=sacsarlm(reg.eq1, data = GCP.data,tol.solve=1.0e-10, listw1,
type="sac",method="LU")##SARAR model AKA Kelejian-prucha, Cliff-Ord or SAC
s<- summary
s(reg5)
impacts(reg5,listw=listw1)
AIC(reg5)
BIC(reg5)
summary(impacts(reg5,listw=listw1,R=500), zstats=TRUE)#For the p-values
s(reg6)
impacts(reg6,listw=listw1)
summary(impacts(reg6,listw=listw1,R=500), zstats=TRUE)#For the p-values
AIC(reg6)
BIC(reg6)
s(reg7)
impacts(reg7,listw=listw1)
summary(impacts(reg7,listw=listw1,R=500), zstats=TRUE)#For the p-values
AIC(reg7)
```

```
BIC(reg7)
s(reg8)
impacts(reg8,listw=listw1)
summary(impacts(reg8,listw=listw1,R=500), zstats=TRUE)#For the p-values
AIC(reg8)
BIC(reg8)
##Likelihood ratio Tests: Test Model Restrictions...H_0:Restricted Coefficients=0
LR.sarlm(reg7,reg4)#Should MANSKI be restricted to SEM? i.e
LR.sarlm(reg1,reg4)
LR.sarlm(reg7, reg3)##or to SAR
LR.sarlm(reg7,reg2)##or to SLX
LR.sarlm(reg7,reg1)##or to OLS
LR.sarlm(reg4,reg1)##SEM to OLS
LR.sarlm(reg3,reg1)##SAR to OLS
LR.sarlm(reg2,reg1)##SLX to OLS
LR.sarlm(reg5,reg1)#SDEM to OLS
LR.sarlm(reg8,reg1)
LR.sarlm(reg6,reg1)#SDM to OLS
LR.sarlm(reg6,reg2)
LR.sarlm(reg2,reg4)
AIC(reg7,reg2)
BIC(reg7,reg2)
AIC(reg7, reg3)
AIC(reg1,reg4)
AIC(reg2,reg3)
AIC(reg2,reg4)

#FA
cortest.bartlett(cor, n=18)# Bartlettâ  ™s Test of Sphericity, test the null hypothesis that correlation matrix is an identity matrix and not ideal for fa
KMO(cor)#Kaiser â  " Meyer- Olkin
cronbach(x)##Cronbach coefï¬  cient Î±
gcp.mle1<- fa(x, nfactors = 2, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
gcp.mle2<- fa(x, nfactors = 3, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
```

```
gcp.mle3<- fa(x, nfactors = 4, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
gcp.mle4<- fa(x, nfactors = 5, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
gcp.mle5<- fa(x, nfactors = 6, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
anova(gcp.mle1,gcp.mle2,gcp.mle4,gcp.mle3,gcp.mle5)
gcp.mle<- fa(x, nfactors = 4, rotate="varimax", fm = "mle", impute="mean",n.rotations=1)
gcp.mle
gcp.mle$scores
fa.diagram(gcp.mle, main = "Factor Analysis using Maximum Likelihood Method")
sc_mle<-gcp.mle$scores
write.csv(sc_mle, file = "sc_mle.csv")
```

# APPENDIX C: Abstract of publication

International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

https://gssrr.org/index.php/JournalOfBasicAndApplied/index

---

## Spatial Regression of the Gross County Product of Kenya on Induced Latent Variables

Samson Tsuma[a*], Prof. Ali Salim[b], Dr. George Matiri[c], Dr. Justin Obwoge[d]

[a]Egerton University, 39, Gilgil 20116 , Kenya
[b,c,d]Department of Mathematics, Egerton University, 536, Nakuru, 20115, Kenya
[a]Email: kenyatsuma@gmail.com, [b]Email: ali.islam@egerton.ac.ke
[c]Email: gemuwa@gmail.com, [d]Email: justinoke69@gmail.com

**Abstract**

Because of a very shallow study carried out to measure regional economic progress in Kenya, we were prompted to investigate on the role of geographical analysis in economic development. The induction of the Gross County Product (GCP) in 2013 had brought about a new viewpoint of assessing the economic growth pattern of Kenya from a single value of the Gross Domestic Product (GDP) to a disaggregate measure that was inclusive of the contributive efforts from each county. Investigating the spatial dependence of this GCP on latent variables solved the error of model misspecification and proved the spill-over effect of the Kenyan economy at the county levels. The Local Indicator of Spatial Association (LISA) (Moran I test) revealed spatial clustering and the Lagrange Multiplier (LM) Test together with the spatial Hausman test suggested an error model fit. Meanwhile, the likelihood ratio test considered a restricted spatial model more suitable than the nested model. Not only was the economic pattern monitored but also a correct version of the 6 economic blocs of Kenya was developed by use of thematic maps where the counties were geographically classified according to the spatial implication.

***Keywords:*** Gross county product; spatial dependence; thematic maps; Latent variables.

## APPENDIX D: Editors Letter

**<u>This document contains the following information (kindly read them carefully):</u>**

1- Internal and External Evaluation Results.
2- Detailed Publication Instructions.

### 1- Internal and External Evaluation Results.

Note: This paper was evaluated based on two stages; the first stage is the internal evaluation, the second stage is the external peer reviewed evaluation. The following results\ remarks are taken from the original reviewers results.

**Reviewers Results: Internal and External Results**

**PART A:** ***Editorial Office Only***

**SECTION I: Internal Evaluation Results**

| Manuscript information | Yes | No |
|---|---|---|
| Is the research within to the scope of the journal? | X | |
| Is it a full paper submission? | X | |
| Is the language of paper English? | X | |
| Will the paper be of interest to its audience? | X | |
| Has the paper or part of it already been published elsewhere? [Based on Google Search on tile and abstract] | | X |
| **Recommendations: Mark where appropriate.** | | |
| Rejected After Internal Review | | |
| Accepted After Internal Review and Recommended for External Technical Review | X | |

**PART B:** ***Reviewers Only***

**SECTION II: External Evaluation Results**

| Mark (X) where appropriate | YES | NO |
|---|---|---|
| Are the references authoritative and representative? | X | |
| Is the paper interesting or relevant for an international audience? | X | |
| Does the title accurately reflect the content? | X | |
| Is there valuable connection to previously published research in this area? | X | |
| Is the abstract sufficiently concise and informative? | X | |
| Do the keywords provide adequate index entries for this paper? | X | |
| Is the purpose of the paper clearly stated in the introduction? | X | |
| Does the paper achieve its declared purpose? | X | |
| Does the paper show clarity of presentation? | X | |
| Do the figures and tables aid the clarity of the paper? | X | |
| Are the English and syntax of the paper satisfactory? | X | |
| Is the paper concise? (If not, please indicate which parts might be cut?) | X | |
| Does the paper develop a logical argument or a theme? | X | |
| Do the conclusions sensibly follow from the work that is reported? | X | |
| Is the Technical Quality suitable for publication? | X | |
| Does the paper provide suitable Contribution To The Field? | X | |
| Is the overall quality suitable for inclusion in this journal? | X | |

**SECTION III - Recommendations: (*Kindly Mark With An X*)**

| | |
|---|---|
| Accept As Is: | |
| Requires Minor Corrections: | X |
| Requires Moderate Revision: | |
| Requires Major Revision: | |
| Submit To Another Publication Such As: | |
| Reject On Grounds Of (Please Be Specific): | |

**SECTION IV: Additional Comments**

Please add any additional comments (Including comments/suggestions regarding the submission, if any):

Discussion on the results is also not very impressive; more clarifications are needed.
Constraints \limitations of the study are not well defined.
Previous studies need more discussion from the author.

## APPENDIX E: NACOSTI Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **398342** Date of Issue: **25/July/2023**

**RESEARCH LICENSE**

**This is to Certify that Mr.. Samson Tsuma of Egerton University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nakuru on the topic: SPATIAL REGRESSION OF THE GROSS COUNTY PRODUCT OF KENYA ON INDUCED LATENT VARIABLES for the period ending : 25/July/2024.**

License No: **NACOSTI/P/23/27999**

**398342**

Applicant Identification Number

Director General

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

**See overleaf for conditions**

**THE SCIENCE, TECHNOLOGY AND INNOVATION ACT, 2013 (Rev. 2014)**
Legal Notice No. 108: The Science, Technology and Innovation (Research Licensing) Regulations, 2014

**The National Commission for Science, Technology and Innovation**, hereafter referred to as the Commission, was the established under the Science, Technology and Innovation Act 2013 (Revised 2014) herein after referred to as the Act. The objective of the Commission shall be to regulate and assure quality in the science, technology and innovation sector and advise the Government in matters related thereto.

**CONDITIONS OF THE RESEARCH LICENSE**

1. The License is granted subject to provisions of the Constitution of Kenya, the Science, Technology and Innovation Act, and other relevant laws, policies and regulations. Accordingly, the licensee shall adhere to such procedures, standards, code of ethics and guidelines as may be prescribed by regulations made under the Act, or prescribed by provisions of International treaties of which Kenya is a signatory to
2. The research and its related activities as well as outcomes shall be beneficial to the country and shall not in any way;
    i. Endanger national security
    ii. Adversely affect the lives of Kenyans
    iii. Be in contravention of Kenya's international obligations including Biological Weapons Convention (BWC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), Chemical, Biological, Radiological and Nuclear (CBRN).
    iv. Result in exploitation of intellectual property rights of communities in Kenya
    v. Adversely affect the environment
    vi. Adversely affect the rights of communities
    vii. Endanger public safety and national cohesion
    viii. Plagiarize someone else's work
3. The License is valid for the proposed research, location and specified period.
4. The license any rights thereunder are non-transferable
5. The Commission reserves the right to cancel the research at any time during the research period if in the opinion of the Commission the research is not implemented in conformity with the provisions of the Act or any other written law.
6. The Licensee shall inform the relevant County Director of Education, County Commissioner and County Governor before commencement of the research.
7. Excavation, filming, movement, and collection of specimens are subject to further necessary clearance from relevant Government Agencies.
8. The License does not give authority to transfer research materials.
9. The Commission may monitor and evaluate the licensed research project for the purpose of assessing and evaluating compliance with the conditions of the License.
10. The Licensee shall submit one hard copy, and upload a soft copy of their final report (thesis) onto a platform designated by the Commission within one year of completion of the research.
11. The Commission reserves the right to modify the conditions of the License including cancellation without prior notice.
12. Research, findings and information regarding research systems shall be stored or disseminated, utilized or applied in such a manner as may be prescribed by the Commission from time to time.
13. The Licensee shall disclose to the Commission, the relevant Institutional Scientific and Ethical Review Committee, and the relevant national agencies any inventions and discoveries that are of National strategic importance.
14. The Commission shall have powers to acquire from any person the right in, or to, any scientific innovation, invention or patent of strategic importance to the country.
15. Relevant Institutional Scientific and Ethical Review Committee shall monitor and evaluate the research periodically, and make a report of its findings to the Commission for necessary action.

National Commission for Science, Technology and Innovation(NACOSTI),
Off Waiyaki Way, Upper Kabete,
P. O. Box 30623 - 00100 Nairobi, KENYA
Telephone: 020 4007000, 0713788787, 0735404245
E-mail: dg@nacosti.go.ke
Website: www.nacosti.go.ke

## APPENDIX F: Copyright Form


International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied



### IJSBAR COPYRIGHT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to IJSBAR and must accompany any such material in order to be published by IJSBAR . Please read the form carefully.

TITLE OF PAPER INCLUDING ALL CONTENT IN ANY FORM, FORMAT, OR MEDIA (hereinafter,
**" Spatial regression of the Gross County Product of Kenya on induced latent variables "):**

COMPLETE LIST OF AUTHORS:

*Mnazi Samson Tsuma, Prof. Ali Salim Islam, Dr. George Matiri, Dr. Justin Obwoge.*

**COPYRIGHT TRANSFER**

The undersigned hereby assigns to the **International Journal of Sciences: Basic and Applied Research (IJSBAR)** ("IJSBAR ") all rights under copyright that may exist in and to the above Work, any revised or expanded derivative works submitted to IJSBAR by the undersigned based on the Work, and any associated written, audio and/or visual presentations or other enhancements accompanying the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the Work; to the extent the Work incorporates text passages, figures, data or other material from the works of others, the undersigned has obtained any necessary permission. **See Retained Rights, below.**

**AUTHOR RESPONSIBILITIES**

IJSBAR distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that The Work is their own and is original. It is the responsibility of the authors, not IJSBAR , to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

**RETAINED RIGHTS/TERMS AND CONDITIONS**

1. Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.

2. Authors/employers may reproduce or authorize others to reproduce The Work and for the author's personal use or for company or organizational use, provided that the source and any IJSBAR copyright notice are indicated, the copies are not used in any way that implies IJSBAR endorsement of a product or service of any employer, and the copies themselves are not offered for sale.

3. Authors/employers may make limited distribution of all or portions of the Work prior to publication if they inform IJSBAR in advance of the nature and extent of such limited distribution.

4. For all uses not covered by items 2 and 3, authors/employers must request permission from IJSBAR .

5. Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.

## INFORMATION FOR AUTHORS

**IJSBAR Copyright Ownership**

It is the formal policy of IJSBAR to own the copyrights to all copyrightable material in its technical publications and to the individual contributions contained therein, in order to protect the interests of IJSBAR , its authors and their employers, and, at the same time, to facilitate the appropriate re-use of this material by others.

**Author/Employer Rights**

If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, IJSBAR assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

**Reprint/Republication Policy**

IJSBAR requires that the consent of the first-named author and employer be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes.

**GENERAL TERMS THAT ARE PART OF RETAINED RIGHTS/TERMS AND CONDITIONS**

1. The undersigned represents that he/she has the power and authority to make and execute this assignment.

2. The undersigned agrees to indemnify and hold harmless IJSBAR from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

3. In the event the above work is accepted or published by IJSBAR and consequently withdrawn by the author(s), the authors agrees to pay withdrawal fees that are equal to the publication fees, by paying withdrawal fees the foregoing copyright transfer shall become null and void and all materials embodying the Work submitted to IJSBAR will be destroyed.

4. In the event the above work is submitted and then it is accepted for publication by IJSBAR, it cannot be withdrawn by the author(s) unless the authors pay withdrawal fees that are equal to the publication fees, if the authors refuse to pay the withdrawal fees; the IJSBAR have the right to publish the paper in any previous (back) volume from the IJSBAR journal without returning to the authors.

5. it should be clear to the authors that if the journal accept the paper then it will send an acceptance letter to the corresponding author (corresponding author is the submitter of the paper to the IJSBAR journal ), if the authors did not respond and satisfy the requirements including the payment of the publication fees within two months from the date of sending the acceptance letter; then the journal have the right to consider this as refuse to pay the withdrawal fees by the authors and the IJSBAR journal will have the right to publish the paper in any previous (back) volume from the IJSBAR journal without returning to the authors.

6. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.

7. By submitting the publication fee, it is understood that the author has agreed to our terms and conditions which may change from time to time without any notice.

8. Before publishing, author must check whether this journal is accepted by his employer, or any authority he intends to submit his research work. we will not be responsible in this matter.

9. If at any time, due to any legal reason, if the journal stops accepting manuscripts or could not publish already accepted manuscripts, we will have the right to cancel all or any one of the manuscripts without any compensation or returning back any kind of processing cost.

10. The cost covered in the publication fees is only for online publication of a single manuscript.

**Mnazi Samson Tsuma** ______________________ **21st / 08 /2023**

**Author/Authorized Agent For Joint Authors** **Date** (-----------------------------------

-----------------------------------

**PLEASE DIRECT ALL QUESTIONS ABOUT THIS FORM TO:**

**The Managing Editor**

editorIJSBAR@yahoo.com

Biblioteca Universitaria de León (Spain)

Brandeis University (USA)