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.

Received: 7/11/2023
Accepted: 8/16/2023
Published: 8/28/2023

---

* Corresponding author.

1. Introduction

1.1 Background information

In 2019, the Kenya National Bureau of Statistics (KNBS) with the help of the World Bank created the Gross County Product (GCP) through the Kenya Accountable Devolution Program (KADP) to progressively monitor and gave an expounded report for users on 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 compilation was according to the international guidelines for the estimation of regional gross domestic product. The process involved the identification and validation 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 reason for these indicators was to give a picture of the economic structure and relative size of the economy for each county [22].

There were 18 indicators used for the GCP calculation which are in details as shown in the appendix Table 13 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 researchers checked on poverty at provincial levels in 2017 [33]. The use of simple linear models for economic assessment had produced unsatisfactory results that did not consider the geographical and boundary effects. Spatial modeling provided convenient model specification tests for a spatial relationship that was not adequately reflected by the linear model [28]. In this study, the spatial effect of the factors that affect the GCP of Kenya was investigated through spatial regression modeling.

Spatial effects which included spillover and externalities across the economies were chief in spelling out economic growth patterns [10, 41]. Spatial interrelation in the analysis of economic growth if side-lined could have resulted in model misspecification. 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 for current prices. This study developed spatial models for the economic performance at the county levels of Kenya.

It was necessary to check if there was a spatial dependence or clustering through the spatial autocorrelation value [30] e.g. 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. The popular spatial 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) [12].

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 variable 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). Due to the large set of econometric data, researchers were vulnerable to making errors on dimensions 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. Furthermore, county economic progress was majorly considered exclusive by majority of people. The GCP 2019 report had neglected the spatial marginal effects and brought forth shallow results of 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. Currently, there was a huge gap in Kenya on spatial modeling with respect to counties. The February 2019 GCP results were only given as a general report and left without having any further value on them. Though a lot of money was used in generating the report, there had not been any further advantage seen from the results. 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 spatial dependence of the GCP of Kenya on its indicators

1.3.2 Specific objectives

- To conduct a dimension reduction procedure on the GCP indicators using factor analysis.
- To determine the best spatial model for the GCP from the resulting factor scores (latent/induced variables).
- To determine the marginal effect of the induced variables on the GCP.

- To draw thematic maps of the indicators of the Gross County Product of the Kenyan counties.

1.4 Justification

The report from the spatial analysis was crucial for monitoring the progress of the economy county wise. 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. Furthermore, with a large set of economic variables came the error from excessive dimensions and assumptions on linearity during modeling. Thus, a conclusive nature of 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. Robustness checks in these 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. Maps that had spatial information were salient for all human practices. Therefore, the thematic maps in this study were of great benefit to stakeholders. For example, quality education was one of the Sustainable development goals (SDG) which were to be achieved by 2030 and also it was one of the variables that affected the Gross county product that was to be mapped county-wise. 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. 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. 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 to 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. For example, clean water and sanitation was one of the Sustainable development goals and was also one of the indicators that affected the Gross county product.

Therefore, the results of the study gave direction and highlighted the progress on the set development blueprint.

1.5 Assumptions of the study

- The Gross county product was an estimate added at county levels to quantify the relative economic size in Kenya.
- The cross-sectional population data was well defined and therefore all potential bias were eliminated.

2. Literature Review

2.1 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. 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 [45].

In Bangladesh, the analysis 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 & infrastructure factor, and fishing & 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 [3].

An evaluation on the country economic rank by applying factor analysis using International monetary fund (IMF) dataset for 20 countries was conducted in 2015. 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 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 [8].

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 [35].

2.2 Application of spatial econometric 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 [36]. 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 in the study among 93 nations that a country's growth rate was positively influenced by the growth rate of neighboring countries through the spill-over effect at their borders [31]. Similarly, 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. 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.

There was a consideration on 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. a spill-over effect of nitrogen oxides emissions among the neighboring provinces was identified 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 [48].

The Gross Regional Domestic Product (GRDP) of Bruto in Central Java Province, Indonesia in 2017 was modelled using spatial regression. 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) [20].

A spatial beta convergence analysis of the real GDP per capita across Germany and Hungary was discussed and in the analysis, the beta convergence with special regards on spatial processes was examined. 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 [46].

In 2007, a provincial study on spatial relationship of poverty in Kenya for both exogenous and endogenous variables that explained the welfare levels in different areas within provinces was conducted. 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 [33].

There was a suggestion on the existence of spatial lag dependence due to the presence of social and spatial interactions over three periods in a 2014 study. The 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 [18].

A spatial-lag model was engaged 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_x) concentrations through diagnostic testing. The results showed that SO_2 pollution levels had a significant impact on housing prices while NO_x pollution did not [24].

Whittle’s spatial autoregressive lag model (SAR) was popularized and extended by distinguishing models in which the disturbances followed a spatial autoregressive process [15, 20]. The model had been applied widely in research e.g. it described the geo-informational phenomena of the housing prices in small municipalities within Pardubice region in Czech Republic [40]. Also, it estimated the sulfur dioxide air pollution concentrations in Canada and obtained an improvement in the pseudo R squared, log likelihood and reduced mean squared error as compared to the base model [21]. An investigation on the spatial structure of the provincial economic growth and the spatial spillover in China from 1998 to 2008 was done using the SEM to account for the spatial autocorrelation [9]. Also, it was shown using the SDEM model 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 [1]. Other researchers used the SDEM and 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 [44]. The determinants of energy efficiency were also analyzed 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 [19].

An exploration 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 was conducted. The study results revealed that distribution of postgraduates in China showed spatial autocorrelation and non-equilibrium and that postgraduate education positively impacted technological innovation [49]. Also in 2016, the SDM obtained from a theoretical model was used and it captured the technology spillovers from a sample of 107 countries for the period 2000–2011 [47].

The Manski model was capitalized on to estimate the private benefits of native vegetation on rural properties in the state of Victoria, Australia [39]. It also 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 [38].

A research on the SARAR spatial model that estimated 526 observations from a random sample collected was done via in-person interviews and indicated that the rent of a multi-unit dwelling decreased by 0.0239% for every 1% increase in network access distance to the nearest major road in Southern Asia-Rajshahi City, Bangladesh [29]. It was also concluded from a 2016 study that the number of road traffic crashes in a given local government area is affected by the number of road traffic crashes from neighbouring local government areas in Oyo state, Nigeria based on the SARAR model [34].

2.4 Summary of literature critique

Unlike the descriptive summaries of data matrices through factor analysis which had been commonly interpreted for country studies [8] and to the level of districts as in Pakistan [45]; the number of factors influencing the GDP classified for Bangladesh GDP [3] and also in Pakistan districts [45], this study dealt exceptionally with Kenyan counties and went further to find out the latent variables which were used in latent spatial regression equations [35].

Though both a spatial relationship on poverty in Kenya and on economic growth in China at provincial levels was investigated [9, 18, 19, 24, 33, 49], this study developed spatial regression models like in Indonesia [1, 20, 44], for 107 countries [47] , in Australia [38, 39], in Nigeria[34], in Czech republic[40], in Canada[21], in Bangladesh[29] and across Germany and Hungary [46]; but for the economic performance at the county levels of Kenya. Furthermore, we checked on spill overs [31] and using the Likelihood ratio test like the study for China provinces [48] to determine whether to use a restricted form of a spatial regression model.

3. Materials and methods

3.1 The dataset and study area

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:

$$\begin{aligned}
&Gross\ County\ Product\ (GCP) = \\
&Education\ (\boldsymbol{Educ}) + Agriculture, forestry\ and\ fishing\ (\boldsymbol{AgrForFish}) + \\
&Mining\ and\ quarrying\ (\boldsymbol{MinQua}) + Manufacturing\ (\boldsymbol{Manuf}) + \\
&Water\ supply;\ waste\ collection\ (\boldsymbol{WatSuWCol}) + Construction\ (\boldsymbol{Constr}) + \\
&Wholesale\ and\ retail\ trade;\ repair\ of\ motor\ vehicles\ (\boldsymbol{WholRRMV}) + \\
&Transport\ and\ storage\ (\boldsymbol{TranSto}) + Accommodation\ and\ food\ service\ activities\ (\boldsymbol{AccoFodSA}) + \\
&Information\ and\ communication\ (\boldsymbol{Infocom}) + Financial\ and\ insurance\ activities\ (\boldsymbol{FinInsuA}) + \\
&Real\ estate\ activities\ (\boldsymbol{RelEstA}) + Professional, technical\ and\ support\ services\ (\boldsymbol{ProTecSupS}) + \\
&Public\ administration\ and\ Defence\ (\boldsymbol{PubAdminD}) + Electricity\ supply\ (\boldsymbol{ElectSup}) + \\
&Human\ health\ and\ social\ work\ activities\ (\boldsymbol{HumHelSocWA}) + \\
&Other\ service\ activities\ (\boldsymbol{Other\ Services}) + \\
&Financial\ intermediation\ services\ indirectly\ measured\ (\boldsymbol{FinIServM1}). \qquad (1)
\end{aligned}$$

It was important to note that equation (1) was evidently miscalculated due to large number of the predictors and thus produced inconsistent results. Furthermore, the GCP was a geographically oriented and therefore the above relationship produced absurd results.

3.1.1 The study area and the counties of Kenya

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) as expounded in Table 12.

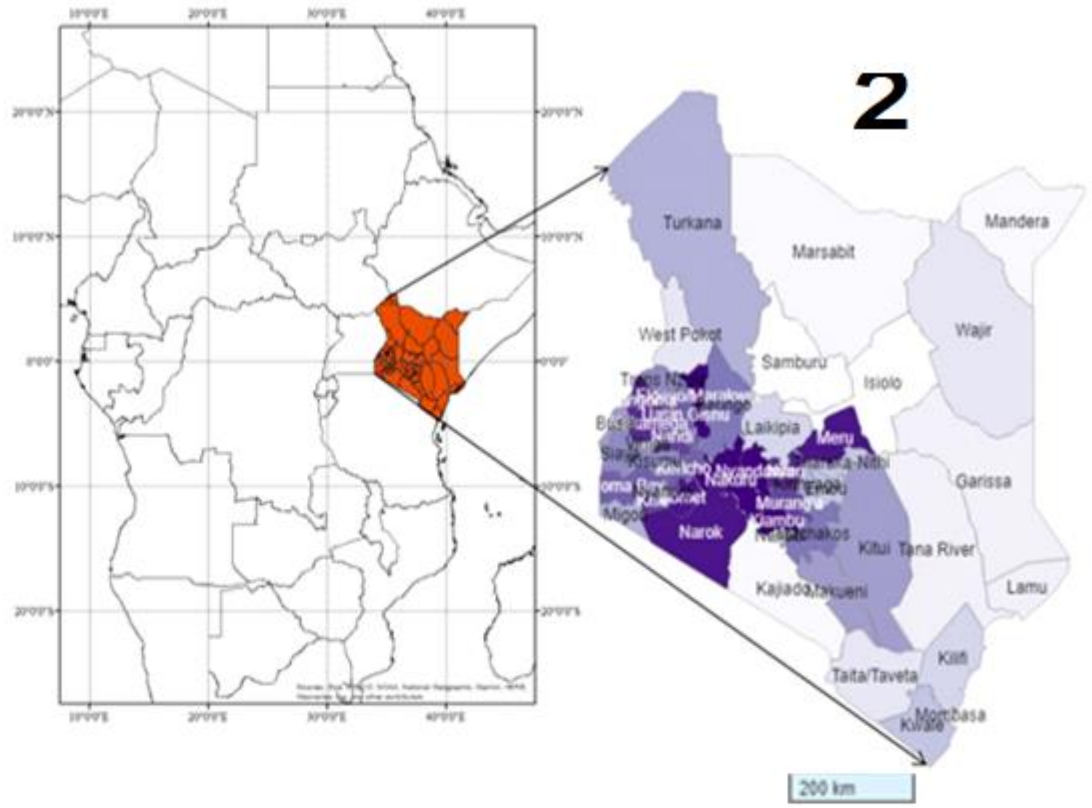


Figure 1: The study area.

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 (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 [42].

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 [4].

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 [37]. 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:

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

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

The test statistic was as follows:

$$Z_h = \frac{I_M - I_{M_O}}{\sqrt{Var(I_M)}}.. \qquad (2)$$

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

Apart from military benefits, it had been proved that mapping was critical for the institutional foundation of economic development and also for civilian and cadastral purposes [50]. 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 $N \times N$ 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. It was shown that when it came to the W matrix, the economic foundation of spatial models was at its weakest. Nevertheless, the conclusion was that the weight matrix had been 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 [14].

The types of contiguity methods that exist include: linear, rook, bishop, double linear, double rook contiguity, and queen. 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 [26].

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 [11].

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 [51].

$$ln\, L(\theta|x) = ln(\prod_{i=1}^{n} f(x_i;\theta)) = \sum_{i=1}^{n} ln\, f(x_i;\theta). \qquad (3)$$

The parameter estimates and fitting of the 7 spatial models was as follows:

3.5.2 Parameter estimation for the simple linear model

The ordinary least squares regression model was given by

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

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

Using the Maximum likelihood technique,

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

With the normality assumption, the MLE was obtained by denoting the likelihood function: $L(\underline{\beta}, \sigma^2)$. Hence the values of β_i and σ^2 maximized $L\left(\underline{\beta}, \sigma^2\right)$.

$$L\left(\underline{\beta}, \sigma^2\right) = f(\underline{y}, \underline{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2} |\sigma^2 \mathrm{I}|^{1/2}} \exp(-\tfrac{1}{2}((Y - X\beta)' \sigma \mathrm{I}(\mathrm{Y} - \mathrm{X}\beta)). \qquad (4)$$

Therefore, the parameters obtained were β_i and σ^2 which were reported in the results.

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

This was given by

$y = X\beta + WX\theta + \varepsilon.$

The parameter estimates by Maximum likelihood estimation were as follows:

$$L\left(\underline{\beta}, \sigma^2\right) = f(\underline{y}, \underline{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2} |\sigma^2 I|^{1/2}} exp\left(-\tfrac{1}{2}(Y - X\beta - WX\theta)' \sigma I (Y - X\beta - WX\theta)\right). \qquad (5)$$

The parameters to be obtained were β_i, σ^2 and θ 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

$y = \rho Wy + X\beta + \varepsilon$.

Using the maximum likelihood function the parameters of the SAR model were estimated as follows.

$$\ln L(\underline{\beta}, \sigma^2) = -{}^{n}/_{2} \ln(2\pi) - {}^{n}/_{2} \ln(\sigma^2) - {}^{1}/_{2\sigma^2} (Y - X\beta - \rho WY)'(Y - X\beta - \rho WY) \ . \qquad (6)$$

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 β_i , σ^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

$y = X\beta + (I_n - \lambda W)^{-1}\varepsilon$.

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

$$lnL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(Z'Z) - \frac{1}{2}(Y - X\beta)'(Z'Z)^{-1}. \qquad (7)$$

Thus, we obtained β_i, σ^2 and λ which was the spatial dependence parameter.

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

This was given by

$y = X\beta + WX\theta + \mu$.

The error term ε was expressed as a function of the vector having auto correlated disturbancesμ.

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

Therefore,

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

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

Substituting in the Log-likelihood function we had;

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

$$[X'(I - \lambda W)'(I - \lambda W)(Y - WX\theta)][X'(I - \lambda W)'(I - \lambda W)X]^{-1} = \hat{\beta}. \qquad (8)$$

The parameters obtained were β_i, σ^2 ,λ which was the spatial dependence parameter and θ values which was a k x 1 parameter vector.

3.5.7 Parameter estimation for the Spatial Durbin Model (SDM)

This was given by

$y = \rho Wy + X\beta + WX\theta + \varepsilon$.

The log likelihood form took the following format;

$$\ln L(\underline{\beta}, \sigma^2) = -{}^{n}\!/_{2} \ln(2\pi) - {}^{n}\!/_{2} \ln(\sigma^2) - {}^{1}\!/_{2\sigma^2} (Y - X\beta - \rho WY - WX\theta)'(Y - X\beta - \rho WY - WX\theta) \qquad (9)$$

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

$y = \rho Wy + X\beta + WX\theta + \mu$.

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

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

$$[X'(I - \lambda W)'(I - \lambda W)(Y(I - \rho W) - WX\theta)]^{-1} = \hat{\beta}. \quad (10)$$

The parameters were β_i, σ^2, λ which was the spatial dependence parameter, $\rho < 1$ parameter value which quantified the spatial dependence of Y on connected regions and θ values which was a k x 1 parameter vector [27].

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

This was given by

$y = \rho Wy + X\beta + \mu.$

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

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;

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

$$X'(I - \lambda W)'(I - \lambda W)(I - \rho W)Y][X'(I - \lambda W)'(I - \lambda W)X]^{-1} = \hat{\beta} \quad (11)$$

The parameters were β_i, σ^2, λ 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 method was used to decide which spatial model was to be preferred to satisfy underlying objectives. The 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 [4]. An original suggestion of the LM/RS test against a spatial error alternative was made and took the form [13].

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

The LM_{err} had an asymptotic $\chi^2(1)$ distribution similar to the LM/RS test against a spatial lag alternative which was given as [6]

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

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

$$\lambda_{LR} = -2[l(\theta_0) - l(\hat{\theta})]. \qquad (12)$$

Based on the ratio of the likelihoods of any two participating models that are nested, the one with the best fit was acquired [23].

For example, if the null hypothesis of the LR test (H0: $\gamma + \delta\beta = 0$) could not be rejected, then the Manski model which was nested could be simplified to the SLX [13] i.e. from equation (12), 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 [12];

- 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
- 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.
- 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 [25].

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;

$$AIC = 2k - 2\ln(\hat{L}). \quad (13)$$

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 [2].

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

$$BIC = kln(n) - 2\ln(\hat{L}). \quad (14)$$

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 [43].

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 Table 1 below summarized the whole analysis process that was implemented in this study.

Table 1: Data analysis guideline

Activity	Software
Data loading and cleaning	R
Exploratory data analysis	R
Factor analysis	R
Shapefile creation	ArcGIS
Spatial models and parameter estimation	R
Model comparison and selection	R
Thematic mapping	R

4. Results

4.1 Exploratory Data Analysis

The variables were presented in the table as follows:

Table 2: GCP indicator variables summary.

Indicator		Min	1st Q.	Median	Mean	3rd Q.	Max	Skew	Kurt.	IQR
AgrForFish	Agriculture, forestry and fishing	1459	19945	47606	60404	78382	301349	2.15	6.04	58436.5
ElectSup	Electricity supply	22	346	581	2994	1310	36932	3.38	11.27	964
MinQua	Mining and quarrying	40	220	620	1244	1444	9643	2.84	9.57	1224.5
Manuf	Manufacturing	11	119	1153	13769	6018	374527	6.02	36.48	5899.5
WatSuWCol	Water supply; waste collection	90	441	740	1191	1142	10819	3.97	16.91	700.5
Constr	Construction	24	1604	3184	9626	6386	175437	5.34	29.88	4783
WholRRMV	Wholesale and retail trade; repair of motor vehicles	1257	3542	5051	13186	7372	294302	6.18	37.96	3830
TranSto	Transport and storage	258	3260	5708	12771	10076	184845	4.8	24.33	6815
AccoFodSA	Accommodation and food service activities	45	166	337	1237	811	14041	3.58	12.29	645
Infocom	Information and communication	143	427	881	2329	1684	53074	6.19	38.08	1257.5
FinInsuA	Financial and insurance activities	260	3876	7380	12897	15414	142765	4.82	26.18	11537
RelEstA	Real estate activities	752	3090	5733	12242	10029	176281	5.35	30.58	6939.5
ProTecSupS	Professional, technical and support services	1	5.5	21	2920.3	228.5	122335	6.39	39.82	223
PubAdminD	Public administration and Defence	2129	4756	5973	7033	7282	40051	4.42	23.01	2525.5
Educ	Education	923	4114	6252	6813	9348	16676	0.58	-0.33	5234
HumHelSocWA	Human health and social work activities	254	1226	2248	2696	3148	17841	3.59	16.17	1923
Other Services	Other Variables	257	1138	1839	1952	2280	8791	2.65	10.56	1141
FinIServM1	Financial Intermediation Services Indirectly Measured	78	375.5	709	5204.6	1497.5	168283	6.82	38.88	1122
TOTAL	Total GCP	15850	72997	103734	160100	173990	1492323	4.7	24.97	100993

The standard normal distribution had a kurtosis of 3 and thus the variables were largely non-normal.

4.1.1 The Correlation plot

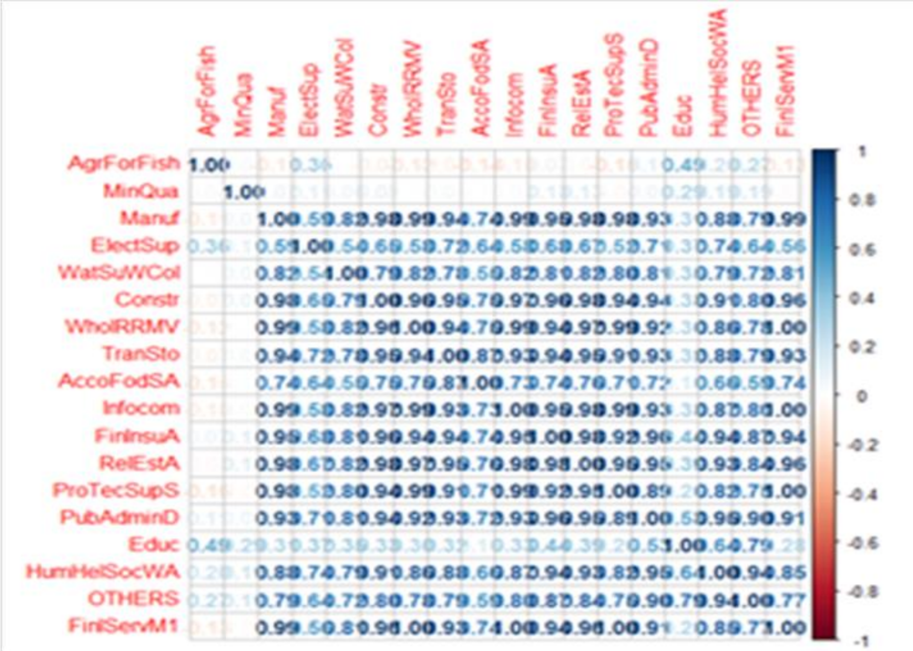


Figure 2: Pearson's correlation plot.

The correlation plot in Figure 2 showed most of the values having a strong positive Pearson's correlation value as presented by the blue 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.

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 as advocated [16], a more accurate recommendation [32] or with the recommended standard value of 0.6 [7] 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. As per KMO measure, a measure of >0.9 is marvellous, >0.8 is meritorious, >0.7 is middling, >0.6 is mediocre, >0.5 is miserable and <0.5 is unacceptable.

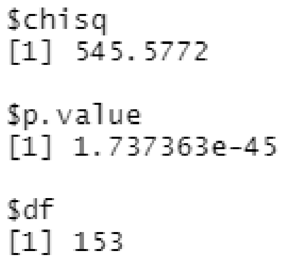
KMO(cor)#Kaiser – Meyer- Olkin

Kaiser-Meyer-Olkin factor adequacy

Call: KMO(r = cor)

Overall MSA (Measure of Sampling Adequacy) = 0.89

The data returned a value sampling adequacy of 0.89 indicating meritorious. Bartlett's test of Sphericity in Figure 3 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 <0.05 indicated that the data did not produce an identity matrix and was thus appropriately multivariate normal and acceptable for FA [17].



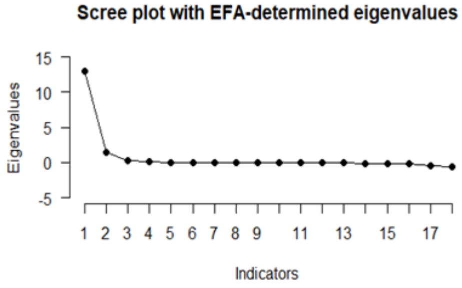


Figure 3: Bartlett's test of Sphericity. **Figure 4:** Exploratory Factor Analysis Scree plot.

The data within this study returned a significance value of 1.737363e-45, indicating that the data was acceptable for FA. The scree plot below in Figure 4 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.

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 3.

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots \ldots . . + l_{1m}F_m + \varepsilon_1.$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots \ldots . . + l_{2m}F_m + \varepsilon_2.$$

$$\vdots$$

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \cdots \ldots . . + l_{pm}F_m + \varepsilon_p . \qquad (15)$$

Table 3: Factor Analysis Model Comparison.

Model	BIC
1	-67.54
2	-148.39
3	-144.12
4	**-159.32**
5	-114.04

From equation (15), model 4 had the deepest negative BIC = -159.32 and thus was suitable for the EFA.

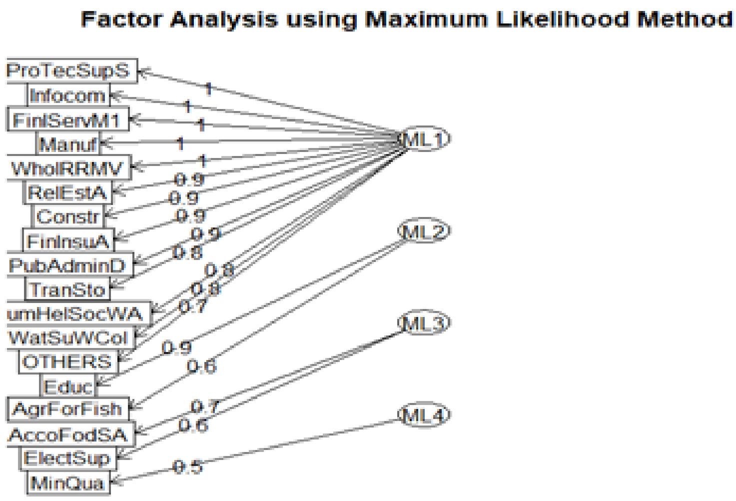


Figure 5: Factor analysis Path diagram.

Based on the path diagram above in Figure 5 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 found after row standardization.

Table 4: Spatial Weights and links.

R code	listw1$neighbours
Number of regions	47
Number of nonzero links	232
Percentage nonzero weights	10.50249
Average number of links	4.93617

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 6.

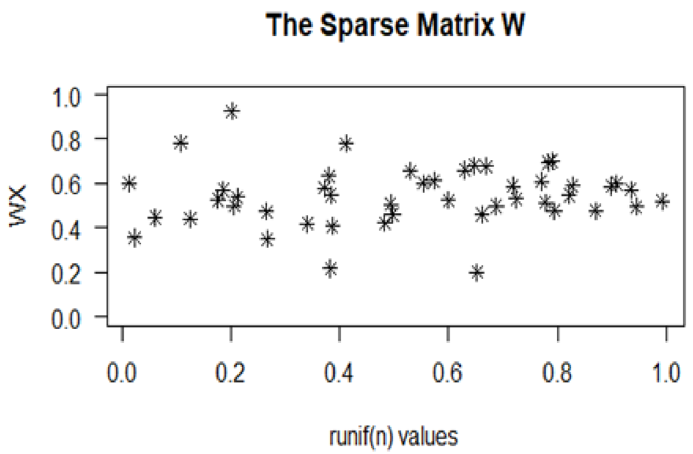


Figure 6: Sparse Matrix W Unmatched Neighbors.

4.1.4 The simple linear model result

Table 5: OLS Results.

R command : reg1<- lm(TOTAL~ ML1+ML2+ML3+ML4, data=GCP.data)				
RESULTS				
Call: lm(formula = TOTAL ~ ML1 + ML2 + ML3 + ML4, data = GCP.data)				
Min	1Q	Median	3Q	Max
-60350	-19108	-5805	7901	158545
Coefficients	Estimate(β_i)	Std. Error(σ^2)	t value	Pr(>|t|)
(Intercept)	160100	6544	24.465	< 2e-16 ***
ML1	193867	7206	26.904	< 2e-16 ***
ML2	30982	7627	4.062	0.000208 ***
ML3	38066	12014	3.168	0.002856 **
ML4	-13477	11485	-1.173	0.247246
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 44860 on 42 degrees of freedom				
Multiple R-squared: 0.9626,		Adjusted R-squared: 0.959		
F-statistic: 270.3 on 4 and 42 DF		p-value: < 0.00000000000000022		

The OLS in Table 5 had a significant result though the ML4 was not needed in the model. The adjusted R-squared value was 0.959 and the model had a p-value of < 0.00000000000000022. According to equation (4), 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 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 as in equation (2).

Table 6: Global Moran I Test.

R command: lm.morantest(reg1,listw1)			
Global Moran I for regression residuals			
model: lm(formula = TOTAL ~ ML1 + ML2 + ML3 + ML4, data = GCP.data)			
weights: listw1			
Moran I statistic standard deviate = 2.8221		p-value = 0.002386	
alternative hypothesis: greater			
Sample estimates	Observed Moran I	Expectation	Variance
	0.211776484	-0.039103244	0.007902968

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 6 was positive 0.211776484 and the p-value = 0.002386 as displayed in the Moran plot below.

The Moran plot was shown below for the residuals;

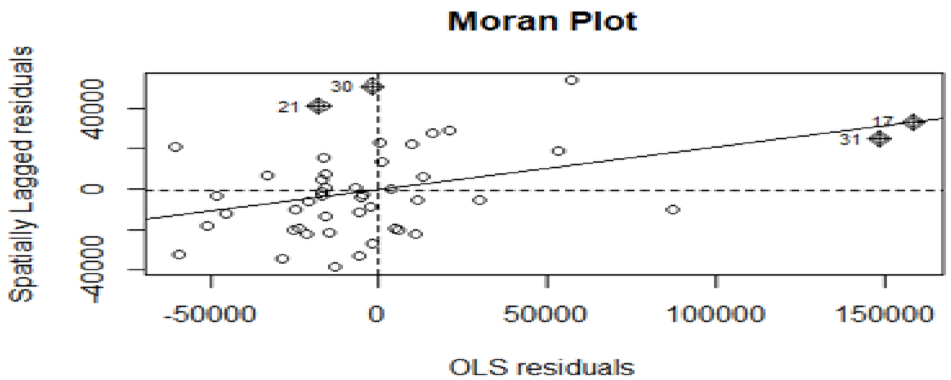


Figure 7: Moran I plot.

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

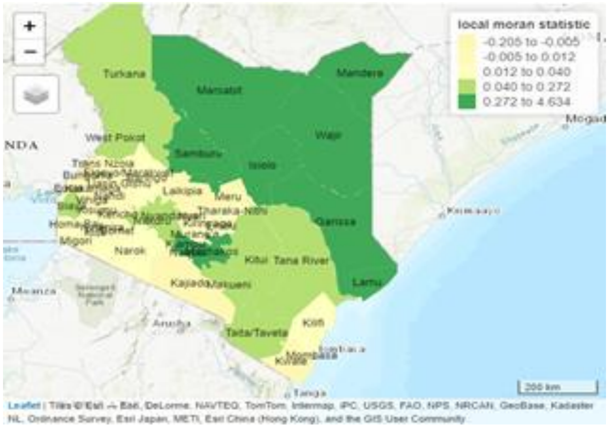


Figure 8: Local Moran Map.

From the Figure 8, it was possible to observe the variations in autocorrelation across space through the Local Indicator of Spatial association (LISA) clustering principle [5]. In the local Moran I statistic, each point each location received its own I value unlike the global Moran I statistic which we got 0.211776484. 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).

From Figure 9, it is apparently clear that there was a statistically significant geographic pattern to the clustering of the Total GCP in Kenya.

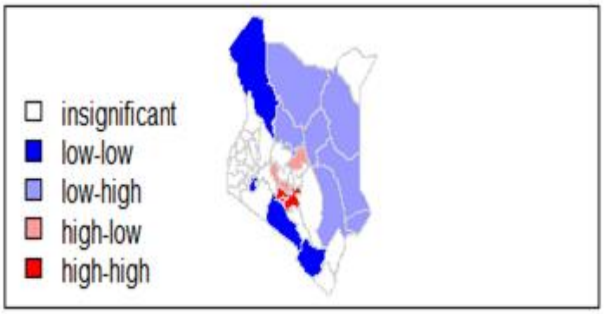


Figure 9: Total GCP Geographical Clusters.

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 7: Lagrange Multiplier/ Rao score Test.

Lagrange multiplier diagnostics for spatial dependence
model: lm(formula = TOTAL ~ ML1 + ML2 + ML3 + ML4, data = GCP.data)
weights: nb2listw(queen.nb)

	Statistic	Parameter	P-value
LMerr	4.68424	1	0.03044 *
Lmlag	3.64167	1	0.05635.
RLMerr	1.81999	1	0.17731
RLMlag	0.77743	1	0.37793
SARMA	5.46166	2	0.06517.

Signif. 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 7 the Lmerr was statistically significant with a p-value 0.03044 * 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 8: Spatial Error Model Results.

R commands	reg4<- errorsarlm(reg.eq1, data=GCP.data,listw1, tol.solve=1.0e-19)##SEM summary(reg4) errorsarlm(formula = reg.eq1, data = GCP.data, listw = listw1, tol.solve = 1e-19)			
Residuals:	Min -61827.2	1Q -18535.0	Median -5476.4	3Q 10087.6 / Max 141932.9
Type: error	Coefficients: (asymptotic standard errors)			
	Estimate(β_i)	Std. Error(σ^2)	z value	Pr(>|z|)
(Intercept)	159809.2	11882.5	13.4491	< 0.00000000000000022*
ML1	192073.2	6940.7	27.6734	< 0.00000000000000022*
ML2	30559.6	7219.7	4.2328	0.00002308*
ML3	38472.3	10265.1	3.7479	0.0001783*
ML4	-22577.3	10562.1	-2.1376	0.0325511*
Lambda (λ): 0.52736	LR test value: 6.0567	Asymptotic standard error: 0.15208		p-value: 0.013854
z-value: 3.4676	p-value: 0.00052511			
Wald statistic: 12.024	p-value: 0.00052511			
Log likelihood: -564.4529 for error model	ML residual variance (sigma squared): 1474500000,(sigma: 38399)	Number of parameters estimated: 7	AIC: 1142.9 (AIC for lm: 1147)	

The SEM had a significant result and all latent variables were needed in the model. The model had a p-value of 0.013854, a lambda value of 0.52736 and an error term ε of 38399. According to equation (7), the model therefore was given as

$$TOTAL = 159809.2+192073.2\ ML1+30559.6\ ML2+38472.3\ ML3-22577.3ML4+(I_{47} - 0.52736W)^{-1}38399.$$

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 μ for auto-correlated disturbances was given as follows;

$\mu = 0.52736W\mu + 38399.$

$\mu(I_{47} - 0.52736W) = 38399.$

$\mu = (I_{47} - 0.52736W)^{-1}38399.$

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 as seen from equation (12).

Table 9: Likelihood Ratio Test.

Model	OLS	SLX	SAR	**SEM**	SDM	SDEM	SARAR	MANSKI
OLS		0.06254	0.04773	0.01385	0.03632	0.04455	0.02681	0.06415
SLX	0.06254		0.1701	**0.4095**	0.08602	0.1194	0.4263	0.2276
SAR	0.04773	0.1701			0.09265	0.114	0.06851	0.1571
SEM	**0.01385**	**0.4095**			**0.2119**	**0.2568**	**0.2771**	**0.3214**
SDM	0.03632	0.08602	0.09265	0.2119			0.1991	0.9078
SDEM	0.04455	0.1194	0.114	0.2568			0.2477	
SARAR	0.02681	0.4263	0.06851	0.2771	0.1991	0.2477		0.3233
MANSKI	0.06415	0.2276	0.1571	0.3214	0.9078		0.3233	

From the Table 9 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 10. Further tests had to prove that the SEM model was better through the AIC/BIC.

4.4.3 AIC and BIC

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

Table 10: Information Criterions.

Model	AIC	BIC
OLS	1146.963	1158.063
SLX	1146.02	1164.521
SAR	1145.043	1157.994
SEM	**1142.906**	**1155.857**
SDM	1145.072	1165.424
SDEM	1145.595	1165.946
SARAR	1143.725	1158.526
MANSKI	1147.059	1169.261

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

4.4.4 Spatial Hausman Test

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

Table 11: Spatial Hausman Test.

Hausman test (asymptotic)	Degrees of freedom	p-value
3.381	5	0.6415

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

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.

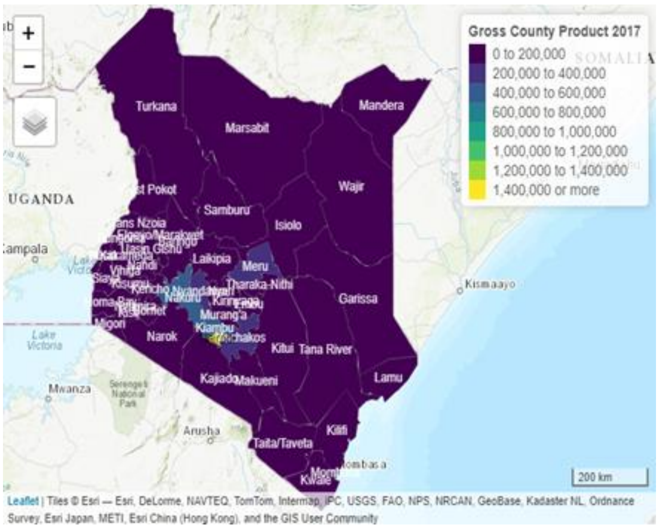


Figure 10: Total Gross County Product Map.

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.

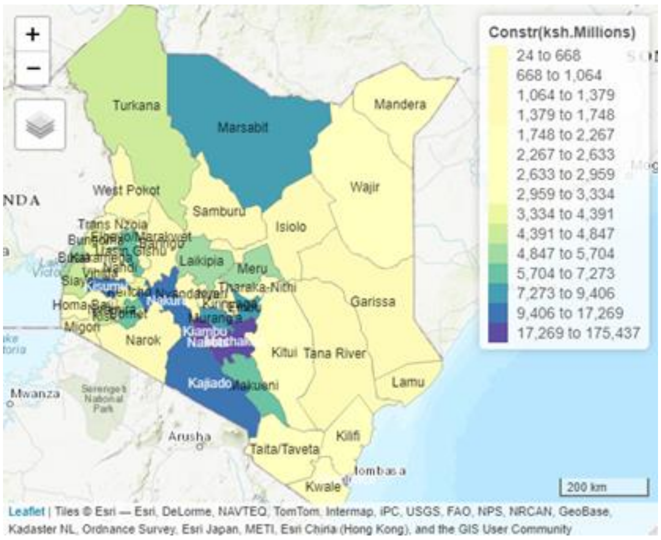


Figure 11: Map of Kenya showing the GCP indicator Construction.

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 12.

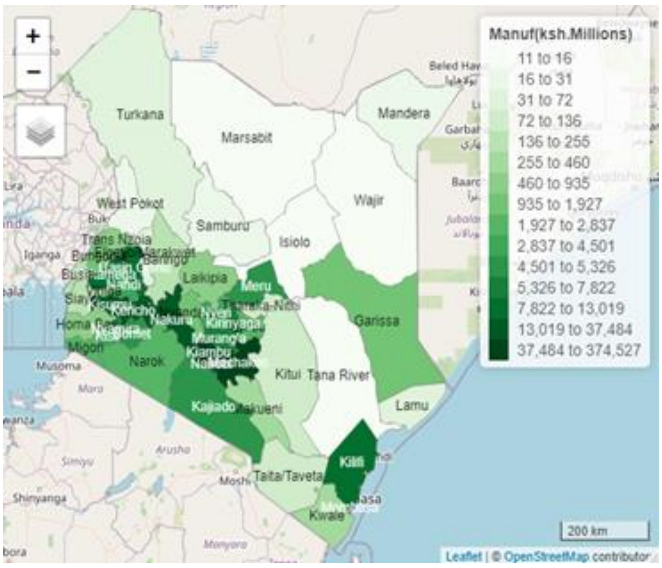


Figure 12: Map of Kenya showing the GCP indicator manufacturing.

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 13.

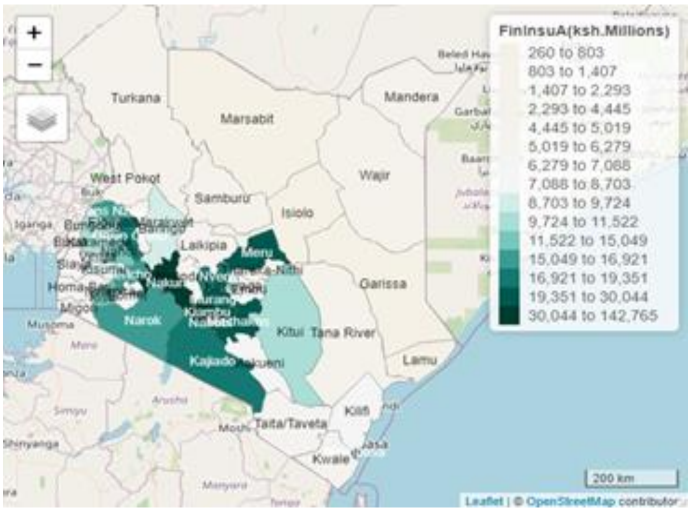


Figure 13: Map of Kenya showing the GCP indicator Financial and Insurance Activities.

ElectSup in Figure 14 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.

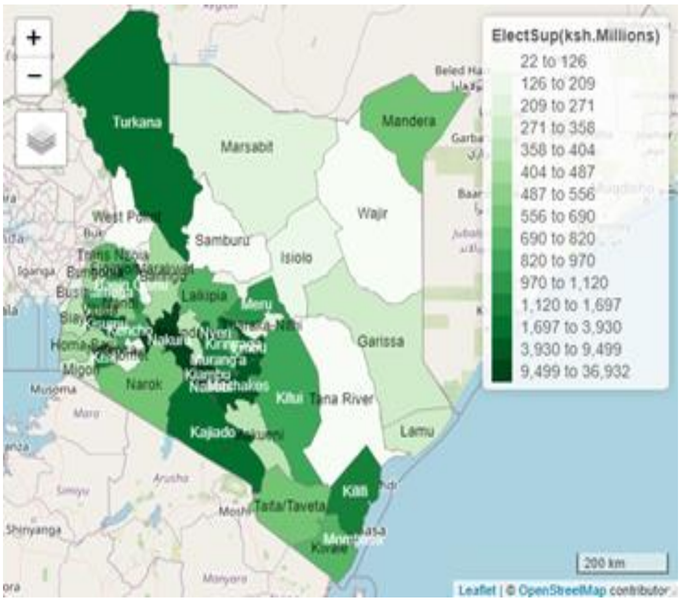


Figure 14: Map of Kenya showing the GCP indicator Electricity Supply.

AgrForFish in Figure 15 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.

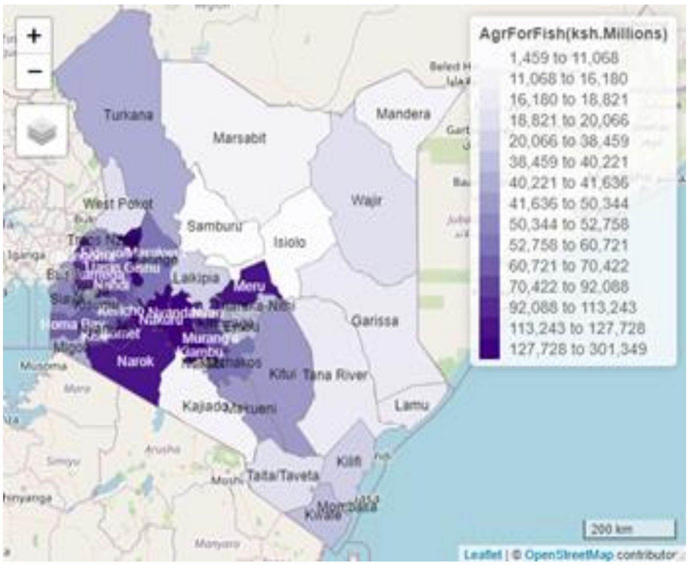


Figure 15: Map of Kenya showing the GCP indicator Agriculture, forestry and fishing.

WatSuWCol in Figure 16 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.

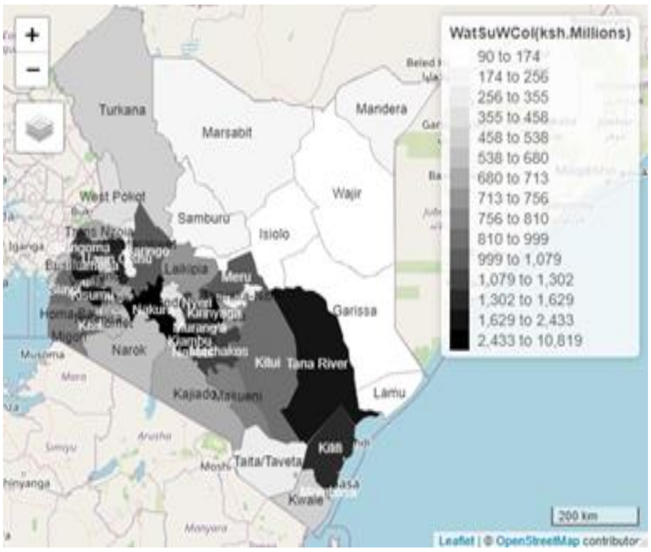


Figure 16: Map of Kenya showing the GCP indicator Water supply and waste collection.

WholRRMV in Figure 17 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.

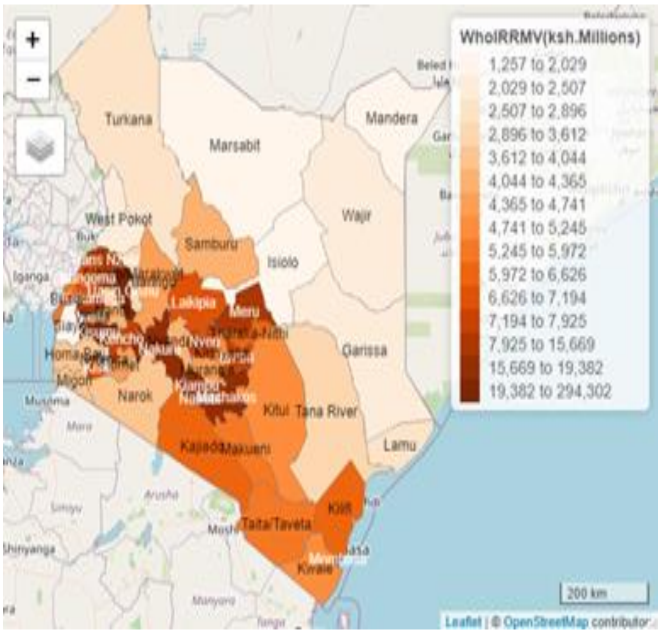


Figure 17: Map of Kenya showing the GCP indicator Wholesale and retail trade; repair of motor vehicles.

Infocom in Figure 18 meant **Information and communication** which included Tele-communications, IT and other Information service activities. Large values were observed in Nairobi, Mombasa and Kiambu counties.

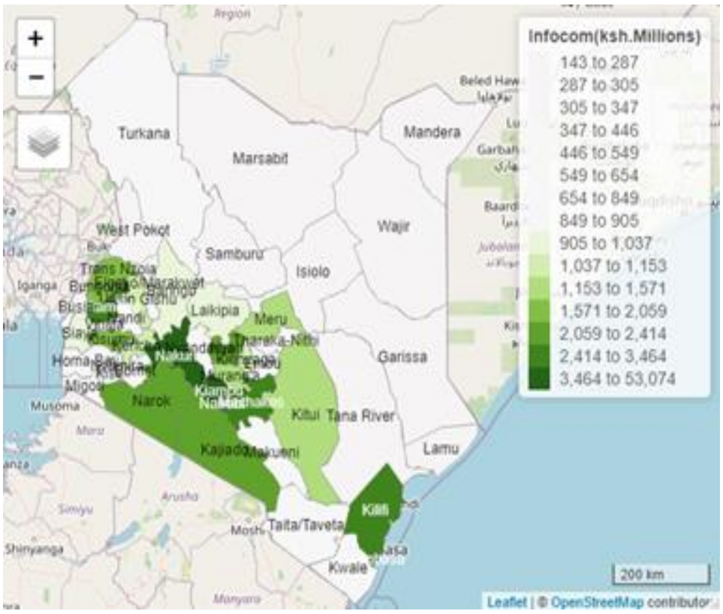


Figure 18: Map of Kenya showing the GCP indicator Information and Communication.

RelEstA in Figure 19 meant Real estate rental activities. Large values were observed at Nairobi, Mombasa and Kiambu counties.

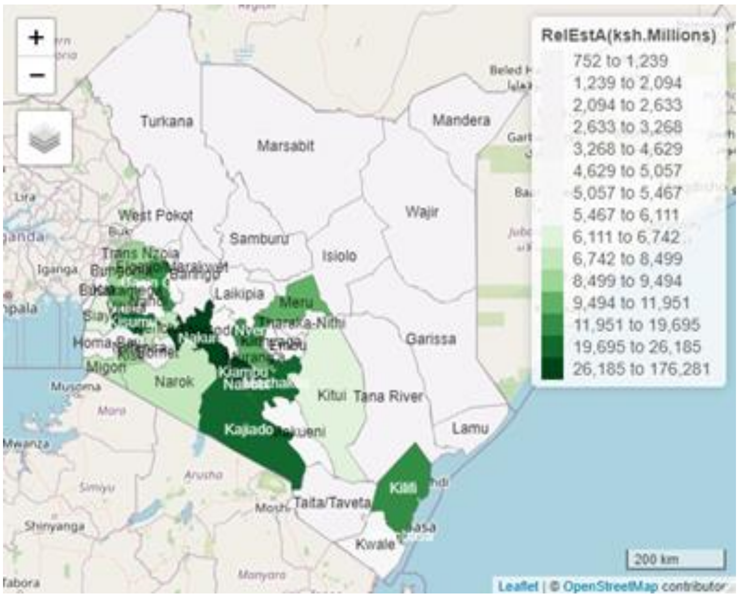


Figure 19: Map of Kenya showing the GCP indicator Real estate rental activities.

PubAdminD in Figure 20 meant **Public administration and Defence** which included compulsory social security. Low values were observed in Marsabit, Lamu, Tana River and Tharaka Nithi.

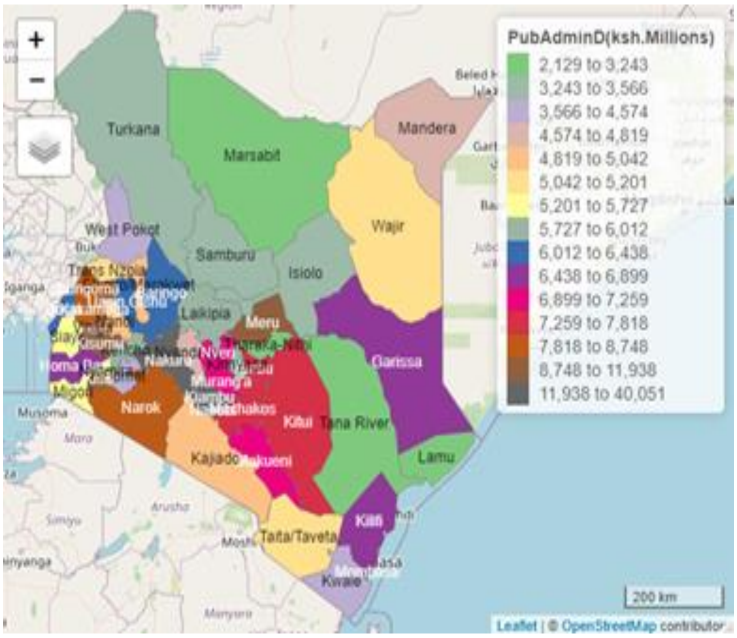


Figure 20: Map of Kenya showing the GCP indicator Public administration and Defence.

AccoFodSA in Figure 21 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.

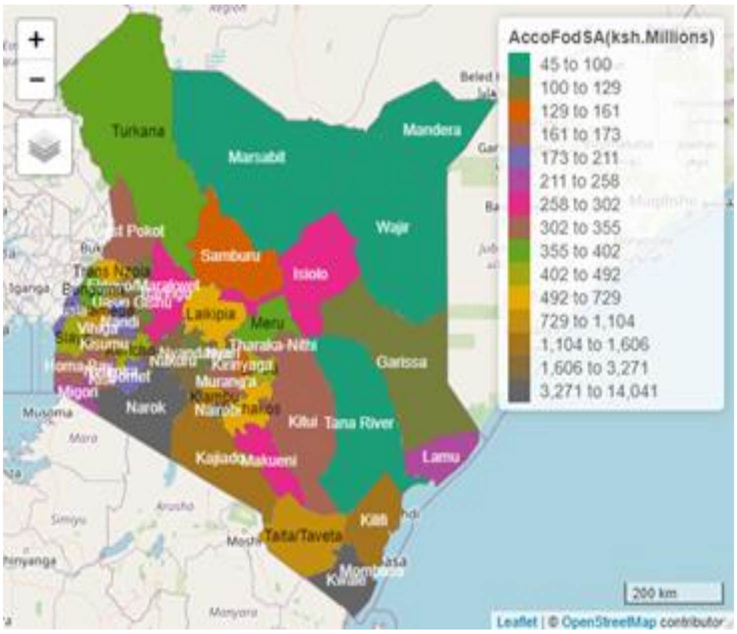


Figure 21: Map of Kenya showing the GCP indicator Accommodation and food service activities.

Educ in Figure 22 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.

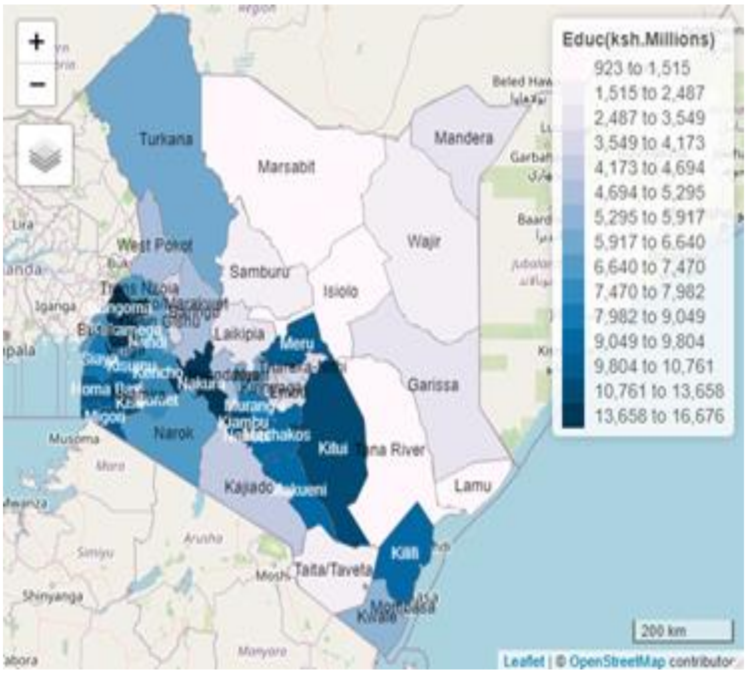


Figure 22: Map of Kenya showing the GCP indicator Education.

TranSto in Figure 23 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.

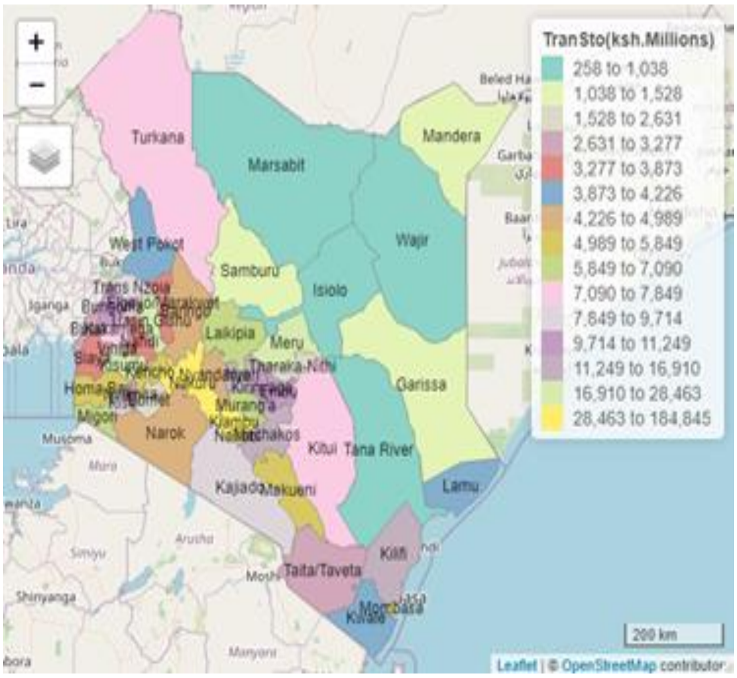


Figure 23: Map of Kenya showing the GCP indicator Transport and storage.

ProTecSupS in Figure 24 meant **Professional, technical and support services** which included Professional, scientific and technical activities. Large values observed in Nairobi and Mombasa Counties.

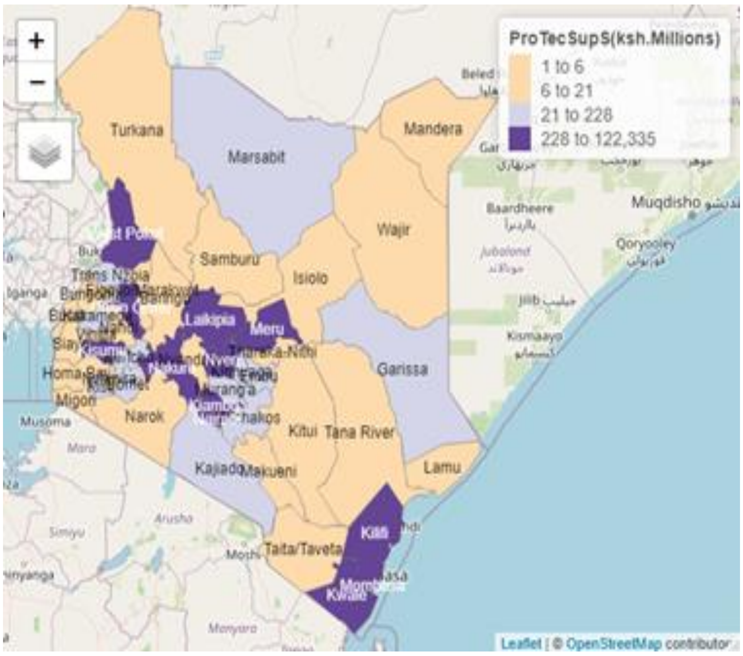


Figure 24: Map of Kenya showing the GCP indicator Professional, technical and support services.

MinQua in Figure 25 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.

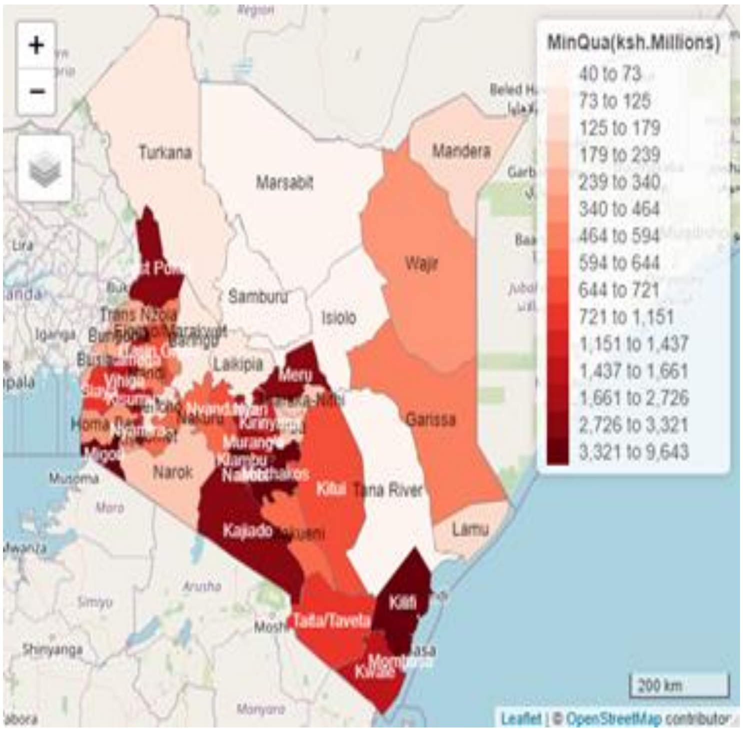


Figure 25: Map of Kenya showing the GCP indicator Mining and quarrying.

HumHelSocWA in Figure 26 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.

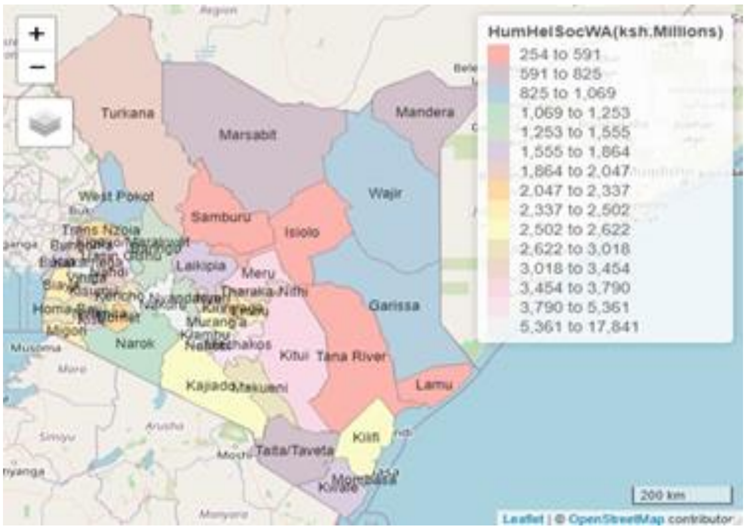


Figure 26: Map of Kenya showing the GCP indicator Human health and social work activities.

Other services in Figure 27 meant other related activities that are not part of the other variables like other entities had larger values in the Nairobi and Nakuru counties.

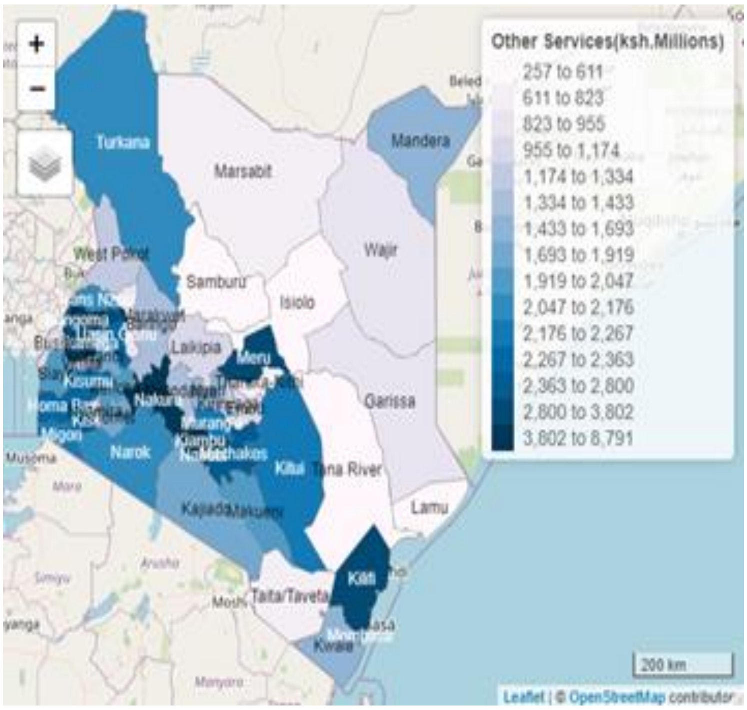


Figure 27: Map of Kenya showing the GCP indicator other service activities.

FinIServM1 in Figure 28 meant **financial intermediation services indirectly measured** which included monetary intermediation which is subtracted. Large values were observed in Nairobi, Mombasa and Uasin Gishu counties.

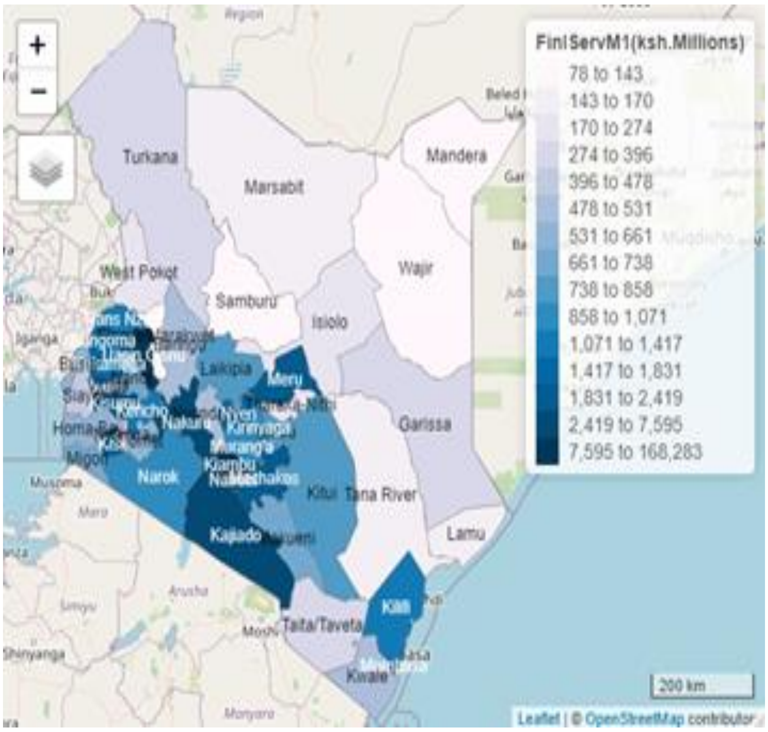


Figure 28: Map of Kenya showing the GCP indicator financial intermediation services indirectly measured.

Table 12: Kenyan Economic blocs' summary.

Economic Bloc(Number of Counties)	**Counties**	**GCP Performance**
Frontier Counties Development Council (FCDC) (7)	Garissa, Wajir, Mandera, Isiolo, Marsabit, Tana River and Lamu	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 and Figure 5 and suggested on how the economic blocs should have been arranged.
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	

5. Conclusion

5.1 Introduction to summary

This chapter presented the summary of the findings, 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. Based on the expenditure approach, the components of the GDP (Y) were four and they include:

- Consumption (C): The total spending by households on goods and services.
- Investment (I): The total spending on goods that were used in the future to produce more goods.
- Government Purchase (G): The spending on the goods and services purchased by government at the federal, state, and local levels.
- Net Exports (NX): This was the exports minus imports.

The GCP was necessary for:

- Estimation of revenue potential for each county.
- Informing economic progress at the county level.
- An indicator for potential for private sector investment.

- 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 2017, 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 objectives of the study were satisfied in the following pattern, from the main objective which was the first to the rest which were specific.

- To investigate the spatial dependence of the Gross county product (GCP) of Kenya on its indicators
- To conduct a dimension reduction procedure on the GCP indicators using factor analysis.
- To determine the best spatial model for the GCP from the resulting factor scores (latent/induced variables).
- To determine the marginal effect of the induced variables on the GCP.
- To draw thematic maps of the indicators of the Gross County Product of the Kenyan counties.

The research approach used was a quantitative, 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 2017 from KNBS was used as the secondary dataset. The data was analyzed in R and ArcGIS and findings presented in the fourth chapter by making use of maps, statistical diagrams, figures and tables. The findings revealed that there was a residual pattern represented as a vector of auto correlated disturbances $\mu = [(I_{47} - 0.52736W)^{-1}38399] \times 10^6$ that was causing the spatial phenomenon that was not only in a single place but among the regions and beyond since the SEM was a global model. 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.

- Main Objective: To investigate the spatial dependence of the Gross county product (GCP) of Kenya on its indicators

This was the main one and the findings of the research was that there was a spatial dependence between the gross county product of Kenya on its indicators and this was how it was found out:

Objective 1: To conduct a dimension reduction procedure on the GCP indicators using factor analysis.

- Since there were a large number of factors totaling to 18, factor analysis reduced them to 4 factors through the maximum likelihood method and the factor scores obtained were then the latent variables for the next procedure.
- The reliability value of our data for factor analysis was the Cronbach coefficient α=0.8861494 which was above the standard value of 0.7.
- 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 545.5772 which was significant with p-value of 1.737363e-45 thus making the factor analysis acceptable.
- The factor models were compared using an ANOVA and the one with 4 factors had the least BIC=-159.32 was chosen.
- Factor scores were obtained from the 4-factor models and labelled as latent variables for the next objective.

Objective 2: To determine the best spatial model for the GCP from the resulting factor scores (latent/induced variables).

- A Moran I test revealed the presence of spatial autocorrelation which was positive at 0.211776484 and with the p-value = 0.002386.
- The Lagrange Multiplier/ Rao score test gave LMerr as statistically significant with a p-value 0.03044 * at an alpha level of 0.05.
- The Spatial Error Model had a p-value of 0.013854, a lambda value of 0.52736 and an error term ε of 38399, BIC value of 1155.857 and AIC value of 1142.906 which were the least among the rest of the spatial models. The model therefore was given as

TOTAL = 159809.2+192073.2 ML1+30559.6 ML2+38472.3 ML3-22577.3ML4+$(I_{47} - 0.52736W)^{-1}38399$.

- The Likelihood Ratio test showed that the SEM was not to be restricted to simpler forms according to Table 8 and the Spatial Hausman test gave a value of 3.381(p-value=0.6415) which was not significant thus no endogenous regressors.

Objective 3: To determine the marginal effect of the induced variables on the GCP.

- 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.

Objective 4: To draw thematic maps of the indicators of the Gross County Product of the Kenyan counties

Each of the 18 indicators was mapped through R software tmap package and the following was observed.

- Large values of the GCP were concentrated in the Nairobi, Kiambu, Nakuru and Mombasa Counties.

- Large values of the Construction indicator were in Nairobi and Kiambu counties
- High values of Manufacturing were in Nairobi and Kiambu counties
- High values of Financial and insurance activities were observed in Nairobi and Kiambu Counties
- Low values of Electricity supply were observed in Tana River, Wajir, West Pokot and Samburu counties
- Large values of Agriculture, forestry and fishing in Nakuru, Nyandarua, Kiambu and Elgeiyo Marakwet.
- Large values of Water supply and waste collection at Nairobi and Kirinyaga counties.
- Large values of Wholesale and retail trade; repair of motor vehicles were observed in Nairobi and Mombasa counties.
- Large values of Information and communication were observed in Nairobi, Mombasa and Kiambu counties.
- Large values of Real estate rental activities were observed at Nairobi, Mombasa and Kiambu counties.
- Low values of Public administration and Defence were observed in Marsabit, Lamu, Tana River and Tharaka Nithi.
- Low values of Accommodation and food service activities were observed in Marsabit, Mandera, Wajir and Tana River counties.
- Large values of Education were found in Kakamega and Bungoma counties.
- Small values of Transport and storage were observed in Marsabit, Wajir, Isiolo and Tana River counties.
- Large values of Professional, technical and support services observed in Nairobi and Mombasa Counties.
- Large values of Mining and quarrying found in Machakos, Kiambu, Kilifi, Migori, Meru and West Pokot.

- Low values of Human health and social work activities were observed in Samburu, Isiolo, Tana River and Lamu counties.
- Large values of financial intermediation services indirectly measured were observed in Nairobi, Mombasa and Uasin Gishu counties.
- Other service activities that are not part of the other variables had larger values in the Nairobi and Nakuru counties.
- The current Kenyan economic blocs are biased and need to be Figure 8.

5.3 Conclusion

All of the objectives were well satisfied and the analysis of the GCP using spatial models had brought about new opinions as shown in the summary about the economic performance of the regions of Kenya.

5.4 Recommendations

Further research was recommended on the;

- Use of other proxies to define the regional spatial economic model in Kenya.
- Use of spatial temporal models
- Continuous production of regional economic data annually to monitor and produce spatial econometric reports.
- For the stakeholders, most of the effects of the Gross County Product was in Nairobi and there should be a way of stretching the benefits to all other 46 regions to avoid economic biases. This could be done by creating opportunities in the other regions.
- Need for county governments to seek the wellbeing of their economy by doing detailed economic research.

5.5 Limitations

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.

5.6 Summary of the chapter

The findings of the study were that there was a vector of auto correlated disturbances that caused a spatial phenomenon in the 2019 geospatial dataset that was in the Kenyan economy report of 2019 which had not been observed by the researchers.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Ali Salim Islam for their invaluable guidance constant support and constructive feedback throughout the entire duration of this research project. Their expertise and passion for the subject matter greatly influenced the quality and direction of this thesis. I am also grateful to the members of my thesis committee Dr. Justin Obwoge and Dr. George Matiri for their insightful input and fruitful discussions during the various stages of this research. Their 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.

References

[1] 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.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions onautomatic control*, *19*(6), 716-723.

[3] 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.

[4] Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in asingle cross-ection. *Papers in Regional Science*, *92*(1), 3-17.

[5] 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.

[6] Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, *20*(1), 1-17.

[7] Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science*, *16*, 74-94.

[8] 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.

[9] Bai, C. E., Ma, H., & Pan, W. (2012). Spatial spillover and regional economic growth in China. *China Economic Review*, *23*(4), 982-990.

[10] Balash, V., Balash, O., Faizliev, A., & Chistopolskaya, E. (2020). Economic growth patterns: Spatial econometric analysis for Russian regions. *Information*, *11*(6), 289.

[11] Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R series*. (*Vol. 747248717*, pp. 237-268). Springer.

[12] Burkey, M. L. (2018). Spatial econometrics and GIS YouTube playlist. *Region 5*(3), R13- R18.

[13] Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A*. *13*(7), 795-800.

[14] Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of*

Regional Science, 52(2), 210-239.

[15] 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.

[16] Cronbach, L.J.(1951).Coefficient alpha and the internal structure of tests. *psychometrika, 16*(3), 297-334.

[17] George, D., & Mallery, P. (2003). SPSS for Windows Sep by Step: A Simple Guide and Reference, Boston, MA: Allyn & Bacon. *Obtenido*

[18] Huang, Y. (2014). Drivers of rising global energy demand: The importance of spatial lag and error dependence. *Energy*, *76*, 254-263.

[19] 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.

[20] 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.

[21] 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 autoregressive model. *Atmospheric Environment*, *79,* 421-427.

[22] 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.

[23] King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference*. Cambridge university press.

[24] 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.

[25] LeSage, J. P., & Pace, R. K. (2009). Spatial econometric models. In *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 355-376). Springer Berlin Heidelberg.

[26] LeSage, J.P. (2015). Theory and practice of spatial econometrics. *Spatial Economic Analysis, 10,* 400-400.

[27] Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, *60*(3), 531-542.

[28] McMillen, D. P. (2012). Perspectives on spatial econometrics: linear smoothing with structured models. *Journal of Regional Science*, *52*(2), 192-209.

[29] 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.

[30] Moran, P.A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

[31] Moreno, R., & Trehan, B. (1997). Location and the growth of nations. *Journal of economic growth*, *2*(4), 399-418.

[32] Nunnally, J. C., & Bernstein, IH (1994). The assessment of reliability. Psychometric theory . 3 (1), 248-292.

[33] 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.

[34] 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.

[35] Oud, J. H., & Folmer, H. (2008). A structural equation approach to models with spatial dependence. *Geographical analysis*, *40*(2), 152-166.

[36] Paelinck, J. H. P. (1978). Spatial econometrics. *Economics Letters*, 1(1), 59-63.

[37] Páez, A., Le Gallo, J., Buliung, R. N., & Dall'Erba, S. (2010). Progress in spatial analysis: introduction. In *Progress in Spatial Analysis* (pp. 1-13). Springer Berlin Heidelberg.

[38] 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.

[39] 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.

[40] 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.

[41] Ramirez, M.T., & Loboguerrero, A.M. (2002). Spatial dependence and economic growth: evidence

from a panel of countries. *Borradores de economia working paper*, 206.

[42] Revelle, W. (2016). How to: Use the psych package for factor analysis and data reduction. *Evanston, IL: Northwestern university, Department of psychology*. https://randilgarcia.github.io/smith-r-workshop/factor.pdf

[43] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

[44] Septiawan, A. R., Handajani, S. S., & Martini, T. S. (2018, May). Spatial Durbin error model for human development index in Province of Central Java. In *Journal of Physics: Conference Series* (Vol. 1025, No. 1, p. 012107). IOP Publishing.

[45] 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.

[46] Szendi, D. (2016) Spatial beta convergence analysis of the real GDP per capita across Germany and Hungary. *Doktornduszok Fóruma Szekciókiadván.* ResearchGate.

[47] Tientao, A., Legros, D., & Pichery, M. C. (2016). Technology spillover and TFP growth: A spatial Durbin model. *International Economics*, *145*, 21-31.

[48] 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.

[49] Xiao, H., & Mao, J. (2021). Effects of postgraduate education on technological innovation: A study based on the spatial Durbin model. *Asia pacific education review*, *22*(1), 89-99.

[50] Yoo, D. (2016, June). Mapping and economic development: Spatial information matters. In *Journal of economic history* (Vol. 76, No. 2, pp. 640-641). Cambridge university press.

[51] Zivot, E. (2009). Maximum likelihood estimation. *Lecture notes on course" Econometric Theory I:Estimation and inference (first quarter, second year PhD)"*.University of Washington.http://faculty.washington.edu/ezivot/econ583/mleLectures.pdf

6. Appendix

Table 13: Variables Table.

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 employees, restaurants, cafes, food kiosks and others.	Explanatory
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