RESEARCH ARTICLE

Inequality and migration in Kenya: Investigating the subnational associations using census data [version 1; peer review: 1 approved with reservations]

Mary Muyonga , Alfred Otieno, George Odipo

Department of Economics, Population and Development Studies, Faculty of Social Sciences, University of Nairobi, Nairobi, Kenya

V1 First published: 26 Nov 2021, 10:1208
<https://doi.org/10.12688/f1000research.74058.1>
 Latest published: 26 Nov 2021, 10:1208
<https://doi.org/10.12688/f1000research.74058.1>

Abstract

Background: Since the early 2000s, there has been an extensive debate on whether migration and inequality are interlinked, with varying conclusions arising from methodological as well as theoretical dispositions. The aim of this study is to contribute to this debate by exploring the nexus between several dimensions of inequalities and migration in Kenya.

Methods: This study used the subnational(county) data on inequalities and migration in Kenya obtained from several reports. Four explanatory variables including access to water, electricity, composite index of County Human Development Index (County HDI) and County Gini were used. Our dependent variable was migration intensity, measured by the Revised Weighted Net Migration Rate. Correlation and spatial regression analysis were performed to model the relationship between migration and inequality.

Results: Findings revealed that migration had a non-linear relationship with income inequality, such that a change in one unit of migration intensity results in a 567 negative change in County Gini. The County Gini had the highest explanatory power in our model, although counties with high HDI tend to have higher migration intensities. Migration intensities in the country were not randomly distributed as we found evidence of spatial clustering with two key emergent hotspots, a high-high in the lake region and a low-low in the coastal region. Regions with low migration intensities correspond with higher poverty, implying that structural factors may explain the migration intensities in the country.

Conclusions: The study highlights that the subnational income inequality reduces as migration intensifies. We conclude that migration has an equalizing effect on inequality as observed in some studies. Regions with high poverty tend to have lower migration intensity, implying that structural factors are important in influencing migration. Use of migration intensity and application of spatial analysis have improved our understanding of migration and

Open Peer Review

Approval Status ?

	1
version 1	?
26 Nov 2021	view
1. Abel Nzabona , Makerere University, Kampala, Uganda	

Any reports and responses or comments on the article can be found at the end of the article.

inequality, and should be applied in future research.

Keywords

migration, inequality, subnational migration, migration intensity, spatial analysis, Geographically Weighted Regression, Kenya

12/11/2021

This article is included in the **Human Migration**

Research

gateway.

Corresponding author: Mary Muyonga (mmuyonga@students.uonbi.ac.ke)

Author roles: **Muyonga M:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Otieno A:** Methodology, Supervision, Validation; **Odipo G:** Methodology, Supervision, Validation

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Muyonga M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Muyonga M, Otieno A and Odipo G. **Inequality and migration in Kenya: Investigating the subnational associations using census data [version 1; peer review: 1 approved with reservations]** F1000Research 2021, 10:1208 <https://doi.org/10.12688/f1000research.74058.1>

First published: 26 Nov 2021, 10:1208 <https://doi.org/10.12688/f1000research.74058.1>

Introduction

One of the emerging research interests in the wake of the Sustainable Development Goals¹ (SDGs) is to understand the nexus between migration and development outcomes. Goal 10 of the SDGs is specifically focused on addressing inequalities among countries, including those related to representations, migration, and development assistance. The importance of inequalities is captured in a United Nations Development Programme (UNDP) report that states that progress on the Millennium Development Goals was hampered by 'unequal access to resources and distribution of power within and among countries' (UNDP, 2005: 52). The growing inequalities within and between countries remains an important discourse in policy circles including the UN Commission on Population and Development Forty-Seventh Session (UN, 2013). Inequality is a multidimensional phenomenon and attracts interests of different disciplines including migration researchers (see Black *et al.*, 2005)². Inequalities, defined as the variations in wellbeing between people or groups of people, becomes an important explanatory factor in migration decision making process.

Migration and inequality linkages are difficult to establish, although scholars have made attempts to understand such linkages. Black *et al.* (2005) observe that the relationship between migration and inequality is governed by access - who gets to migrate, where they migrate to - and the different opportunities that different types of migration streams offer. Whilst their study showed that the migration-inequality relationships vary across space and within and between regions, they also highlight the need to define both the types of migration and types of inequality being analyzed, as the different types of migration may have different effects on different dimensions of inequality. There are economic and social policy constraints affecting migrants as they move for better livelihoods as noted by Klugman (2009), depicting the complexity of human mobility and development. On one hand, inequalities experienced in household settings are a reflection and amplify the constrained opportunity structure (Melamed and Samman, 2013). On the other hand, while migration benefits the family members left behind, it can exacerbate inequalities as migrants tend to come from the better-off backgrounds to start with.

Global scholarly work on migration and inequality as a causal relationship has led to inconclusive results, as some argue that inequality is a prerequisite for migration, while others hold that migration causes inequality. There is consensus that migration is a process that leads to social transformation resulting in changing social structures and creation of new social institutions (UNDP, 2009; de Haas, 2010; King, 2012). When it relates to inequality, migration results in increased opportunities, inequalities and at times, increased poverty, prompting de Haas to opine that 'to understand society is to understand migration, and to understand migration is to better understand society' (de Haas, 2014:16).

Scholars exploring the complex relationship between migration and inequality have mostly relied on econometric analysis of the effects that remittances, as a proxy measure that captures the contribution of migrants, and how that affects the household income and wealth status in the sending communities. Such remittances, especially in monetary form, improve the migrant receiving household welfare and could in fact improve their wealth status in the community. There are numerous studies on the effect that remittances have on income inequality globally (de Haas 2007, 2009; Ebeke and Le Goff, 2011), and a growing body of evidence from country case studies in Africa including Botswana, Burkina Faso, Egypt, Ethiopia, Ghana, Kenya, Nigeria and Somaliland (Plaza, *et al.*, 2011; Muyonga, Odipo and Agwanda, 2020). The results of such studies have yielded conflicting findings, as some indicate that increased migration leads to higher inequality as remittances from migrants reduce income inequality between migrant and non-migrant households in areas of origin; while others find that remittances from migrants increase inequality in areas of destination. The conflicting findings in the effect that remittances have on income inequality are largely based on the methodological approach adopted in the respective studies (Adams, *et al.*, 2008). The econometric approach is not without criticism, the method relies on point estimates of migration event³ which largely ignore the repeat movements across the life cycle, hence ignores the migration system. de Haas (2010) offers a framework for understanding the migration and inequality nexus, pointing out that migration is a social process, a normal process that occurs as societies develop, and inequality is an outcome of that process. Moreover, as societies change, migration processes transition into various forms, meaning that the effect of inequality may also vary. Thus, determination of the effects of migration on inequality requires that we consider migration as part of broader social process (Castels, 2010).

¹Sustainable Development Goals are an Agenda with a specified Action Plan made of 17 Goals and 169 targets aimed at transforming the world by 2030. The SDGs build on to the Millennium Development Goals that were not achieved.

²Most studies specifically looked at the relationship between international migration and inequality. In fact, the paper makes a strong case for the use of multidimensional analysis of inequality not only focusing on income and wealth.

³Point estimates of migration events are captured through population census, which effectively measure the migration event at a fixed period referenced around the census date. It fails to capture any repeat movements that would have been made in the inter censal period, and thus is designated as a measure of events.

Several studies in Kenya have explored the relationship between migration and inequality. The earlier studies considered inequality as a determinant of migration decision, with [Wakajummah \(1986\)](#) study showing that land inequality influenced the propensity to migrate among young males. In their study, [Knowles and Anker \(1981\)](#) consider the effects of remittances on income inequality, and find a weak effect of urban-rural remittances on income inequality. This led to their conclusion that the migration-inequality relationship depends on some intervening variables, including the educational level and income of the migrant, urban or rural residence of the migrant household, and the migrant household wealth status including assets owned and number of dependents. Different findings were observed by [Hoddinott \(1994\)](#) who finds that remittances increase income inequality between migrant and non-migrant households in rural areas. Reflecting on the perceived relationship between land inequality and the increase in migration propensity observed by [Wakajummah \(1986\)](#), a follow up study using data from the 2009 Kenya Migration Household Survey by the World Bank, shows that household predisposing factors influence migration decision making, and therefore, the effect of migration on inequality will depend on such household mitigating factors, hence land inequality is an outcome of other factors ([Bang, et al., 2016](#)). The observed effects of remittances are not similar for urban and rural areas, as noted by [Oyvat and wa Githinji \(2017\)](#), who found that in urban areas, migration results in the influx of migrant workers who may receive lower wages than natives in the urban areas resulting in increased income inequality in the urban areas; while in rural areas, the remittances received from migrants results in higher incomes for the migrant household resulting in improved economic wellbeing thereby increasing income inequality between migrant and non-migrant households. Thus, their study illuminated the mechanisms through which migration influences income inequality comparing urban and rural areas of Kenya.

The earlier studies conducted in Kenya have featured two main approaches, those that consider inequality as a determinant of migration, and others that consider inequality especially income inequality as an outcome of the migration process, but only in regard to the migrant areas of origin. This means that there are still many unknown aspects in understanding how migration as a process affects inequality. While aspects of the effect of inequalities on migration have been studied, they are limited to household analysis and individual migrant experiences. In addition, the effect of migration is measured using remittances sent by migrants, largely ignoring the effect of the wider process and the impact it has on population flows between sending and receiving areas. Moreover, beyond considering land inequality as a determinant for migration, there is little evidence of other dimensions of inequalities considered in earlier studies. Against this backdrop, our study seeks to investigate how migration as a demographic process is affected by inequalities.

The study sought to explore how the inequalities between counties (sub national administrative regions) in Kenya may be related to internal migration movements. This study builds upon works done on migration and inequality in Kenya but differs from previous studies in several ways. First, the study was not a deterministic but rather an exploratory study about how migration intensities change with shifting inequalities. Second, while previous analysis was based on individual migrant characteristics, the study conducted a macro analysis of county level migration and inequality patterns. Third, the study considered the effect of both income and non-income inequalities on migration. Lastly, the study focused on subnational analysis and adopted spatial analysis techniques to understand the effect of inequalities on migration in Kenya.

Data and methods

Data

Migration data was extracted from the 2009 Kenya Population and Housing Census micro data of the Kenya National Bureau of Statistics (KNBS) following a formal data request. The 2009 Population and Housing Census was conducted in August 2009, with the reference night being August 24/25. Data was collected for a period of one week, on all persons residing in Kenya on the census reference night (see [Republic of Kenya, 2012](#)). The unit of enumeration was the household unit and information were collected on type of household, access to social amenities, and demographic information on the household members including education, age, sex and occupation. Migration information was captured using the following variables: place of birth (P18), place of previous residence (P19), duration of residence (P20) and place of enumeration (P21) ([Republic of Kenya, 2012:4](#)). As information was collected on the persons who migrated at two points in time, at birth and at time of enumeration, there may be undercounts of migration transitions, as repeat movements and mortality of migrants are not captured.

For our study, we use two of the migration variables, namely place of birth (P18) and place of enumeration (P21) to generate information on lifetime migrants. Lifetime migrants were identified as persons whose place of birth was different from the place of enumeration at the time of the census. We derived the lifetime migration data from cross tabulating the place of birth by place of enumeration. The end result was a set of contingency flow for all the 47 counties showing in-migration and out-migration flows.

The next step in our analysis was generation of the migration intensity measure. Our dependent variable was migration intensity, a measure that captures both migration rates and impacts (Van Imhoff and Keilman, 1991; Rees *et al.*, 2000; Bell *et al.*, 2002; Liu, *et al.*, 2011; Shi *et al.*, 2020).

For each county, we generated net migration rates using the following formula

$$NM_{ci} = (I_i - O_i) / P_i$$

where NM_{ci} is the net migration for County i , I is the number of in migrants, O is the number of outmigrants and P is the enumerated population

The Revised weighted migration rates, RNM_i was calculated using the formula

$$= (NM_{ci} * (I_{ci} / \sum I_n) * N)$$

where NM_{ci} is the net migration rate for County i , I_{ci} is the number of in migrants in County i , I_n is the total number of migrants in all the 47 counties, and N is the number of counties, in our case, they are 47.

RNM_i considers the proportion of migrants in the total population of a given area, therefore considering the undercounts or overcounts that would occur due to huge differences in total population. When in-migrants are larger than out-migrants, the RNM_i gives a positive result while negative results imply that more migrants are leaving the area. Thus, the RNM_i gives a useful indicator of the intensity of migration and the impact that has had on population redistribution in a given county. Thus, the Revised Weighted Net Migration Rate, gives us the proportional distribution of migrants in a given county as a weighted count of all in-migrants in the country.

Inequality was captured using four key indicators, namely the County Human Development Index (HDI), the County Gini, the proportional access to water and proportional access to electricity within the counties. The four indicators were readily available in different reports based on the 2009 Kenya Population and Housing Census data. The County Gini variable was extracted from the Kenya Inequality Study⁴ jointly conducted by the Kenya National Bureau of Statistics and Society for International Development using a hybrid dataset that combines household data from the 2009 Kenya Population and Housing Census with livelihood data from the 2005 Kenya Integrated Household Budget Survey data to estimate monetary and non-monetary measures of inequality in Kenya (KNBS and Society for International Development, 2013).

The County Gini was derived using the small area estimation technique that followed three key steps. First, data for 1999 Kenya Population and Housing Census was matched to the 2009 Kenya Population and Housing census through a process of matching the clusters of the enumeration areas in these censuses. The use of 1999 census data was necessitated so as to trace the Kenya Integrated Household Budget Survey household clusters, as these were based on the sampling frame used in the earlier census of 1999. The variables that are similar in the 2005/6 KIHBS and 2009 census were identified. Second, a regression model was applied to identify household characteristics and the comparable consumption patterns from the KIHBS survey data. The resultant regression equation was then used to estimate the daily consumption and expenditure patterns using the 2009 variables including household size and other characteristics (KNBS and SID, 2013:3). Finally, through a simulation process, household expenditures for the 2009 census households were inputted using the socio-economic variables estimated using the survey data. Thereafter, the Gini coefficient was computed using consumption expenditure values obtained from the small area techniques. The value of Gini ranges from 0 to 1, with 0 implying there is perfect equality in incomes while 1, implies there is perfect inequality in incomes. A summary table capturing the county level Gini Coefficients can be found in the report (see KNBS and SID, 2013:43).

Two non-income measures of inequality were used in the analysis, namely access to safe water sources, and use of electricity for lighting. Data on access to water and access to electricity was accessed from the *Socio-Economic Atlas of Kenya* report, which is based on deeper analysis of data from the 2009 Kenya Population and Housing Census data, and provides subnational analysis of county and sub-location level data (Wiesmann *et al.*, 2014). Access to safe water sources was obtained from the 2009 census questionnaire, in which all households named their source of domestic water. The indicator of access to water captures the number of households with access to one or more water sources which may include piped, borehole, protected wells, protected springs or rainwater. The information is summarized by county giving

⁴The Kenya Inequality Study used the 2009 Kenya Population and Housing Census data and the 2005/6 Kenya Integrated Household Budget Survey, to generate estimates of poverty and inequality.

the number and percentage of households with access to safe water (Wiesmann *et al.*, 2014:64). In measuring the access to electricity, the study used data on the proportion of households using electricity as source of lighting. The data was extracted from the Socio-Economic Atlas of Kenya report (Wiesmann *et al.*, 2014:78). We used this variable as a measure of the living standards of the household, as electricity distribution in the country is unequal. The variable captures the proportions of households in a given county who indicated that they use electricity as their main source of lighting.

The fourth variable, the County Human Development Index, County HDI was obtained directly from the 2009 Kenya National Human Development Report (UNDP Kenya, 2010). The report assesses the overall changes in the longer term, based on a composite measure of education and literacy rates, healthy living and access to social amenities, the gross domestic product and estimates of earned income by gender. From this report, we obtained estimates of the county HDI which were indicated by Province and district (see Annex 1.1, UNDP Kenya, 2010:77). To generate the values for the present-day counties, a matching process of district to county was employed and average measures used in situations where several districts made up one county, for example, in Nyanza Province, Bondo district and Siaya district are now part of one county, named Siaya County.

Analytical methods

To determine the interrelationship between migration and inequality, the two key analytical techniques were employed, namely correlation and regression analysis using spatial analysis techniques using ArcGIS. While the normal correlation analysis could help to determine if a relationship exists between migration and inequality, and the strength of the relationship, spatial analysis helps to unveil the patterns of flows and their divergence and connectedness. Spatial analysis derives from Tobler's First Law 'Everything is related to everything else, but near things are more related than distant things' (Tobler 1970: 234). It is a measure of relationship between contiguous spatial units and measures spatial dependence or spatial heterogeneity (Anselin, 1990).

The bivariate correlation analysis was conducted using SPSS 22 software. The outcome of the bivariate correlation analysis is the Pearson product-moment correlation, Pearson r (Pearson, 1909). Bivariate correlation assumes that variables are normally distributed but has been found to perform well when normality is violated or when one of the variables is discrete. The values of Pearson r range from -1 to 1 , where -1 indicates there is perfect negative correlation between variables, 0 indicates there is no correlation between variables, while $+1$ indicates there is a perfect positive correlation between the variables, such that a rise in one variable leads to a rise in the other. The output of correlation analysis includes the Pearson r value and the significance of the correlation (2 tailed). An asterisk denotes that the correlation is significant at the 0.05 level, while double asterisk shows correlation is significant at 0.01 level.

The spatial analysis was done using ArcGIS 10.5 software to derive two measures of autocorrelation, the Global Moran I , and the local Moran also known as the local indicator of spatial autocorrelation (LISA) which conducts cluster and hot spot analysis (Anselin, 1995). The Global Moran I determine whether there may be unique patterns, such as incidences of clustering. The Global Moran's I , tests for spatial randomness, thus testing the null hypothesis that the spatial autocorrelation of a variable is zero. If the null hypothesis is rejected, the variable is said to be spatially autocorrelated (Ord and Getis, 1995). The output of the analysis returns five values: the Moran's Index, Expected Index, Variance, z-score, and p value. The value of Moran's I range from -1 to 1 , with -1 , where the value 1 means there is perfect clustering of similar values, while 0 means there is no autocorrelation, hence any clusters arising are of dissimilar values. Thus, a positive value of Moran's I indicates that the values being analyzed tend to cluster spatially, either as high values clustering together, or low values clustering together. A negative index implies that high values repel each other and tend to be near low values. The results also include spatially generated maps that show clustering of migration patterns, as well as areas with divergent characteristics, thus visually clarifies the effect of migration on inequality. A criticism of Moran's I is that the measure is limited only to the strongest associated locations (Wartenberg, 1985). The local tests for spatial association (LISA) helps in spatial cluster identification and spatial filtering (Tiefelsdorf and Boots, 1995, 1997; Hepple, 1998). The formula for calculating LISA (I_i) is expressed in the works of (Anselin, 2017). The output of LISA is evidence of clusters, where regions with high or low values are identified based on their degree of statistical significance, based on the Getis-Ord statistic, $G_i^*(d)$ (Getis and Ord, 1992). The interpretation of z-scores for the G_i^* statistic is quite different from the interpretation of z-scores in the Global Moran's I . The interpretation of the G_i^* Statistic is that a positive association denotes a clustering of high values, while negative association denotes a clustering of negative values. Comparatively, for the Moran's I , positive value of I_i indicates spatial clustering of similar values while a negative value of I_i indicates a clustering of dissimilar values.

Regression analysis was employed to determine the spatial relationship between the migration and inequality variables using two measures, the Ordinary Least Square regression (OLS) and Geographically Weighted Regression (GWR) tools in ArcGIS. The OLS regression analysis was employed to test if our model is effective in explaining the relationship

between the variables. We conducted the Geographical Weighted Regression (GWR), which considers both geographical differences and spatial relationships in the data being analyzed. Geographically Weighted Regression fits a regression equation for all features in the data set using the dependent and explanatory variables, within similar neighborhoods. GWR builds on the OLS by allowing the relationship between the independent and dependent variables to vary by locality. The key assumption of GWR is that the strength and direction of the relationship between the dependent and independent variables is influenced and can be modified by contextual factors (Fotheringham *et al.*, 2003). Several variables were included in the GWR regression equation. The dependent variable Y is migration intensity is measured by the Revised Weighted Net Migration Rate, that captures the temporal effect of migration on population distribution. The independent variables were the county-based measures of inequality including County Gini, County HDI, proportion of persons with access to power and proportion of population with access to water. The resultant equation was as follows:

$$Y = PW_1 Y_1 + B_1 X_1 + B_2 X_2 + B_3 X_3 + B_4 X_4$$

Where, Y is the dependent variable and measures migration intensity, PWY is the autocorrelation factor, while the independent variables include the county inequality measures namely, County Gini, County HDI, proportion of persons with access to electricity in a given county, and proportion of population with access to water in a given county and B1, B2, ... Bn are the coefficients to be estimated.

The OLS regression analysis yields an output that contains the following information: the OLS residuals, statistical results and diagnostics, a table of explanatory variables and their coefficients (called the OLS Summary Report) and a table of the regression diagnostics⁵. To interpret the results, we focus on the R-squared measures which shows how much change in the model is caused by the dependent variable, in our case, the migration intensity.

While the OLS regression is useful in providing an indication of the model efficacy, it has limitations. First, it does not cater for the spatial effects. Second, when there two or more variables that can affect the dependent variable and also affect each other, OLS regression will not be able to counter this multicollinearity effect, thus, the analysis will show that variables which may otherwise be significant in the analysis, rendered statistically insignificant (Young, 2018; Shrestha, 2020).

As a result, we applied as second regression model, the Geographically Weighted Regression (GWR) to the data, to cater for multicollinearity. The output of the Geographical Weighted Regression Analysis comprises of five features. These include; fields for observed and predicted *response* values, condition number (cond), Local R², explanatory variable coefficients, and standard errors generated by ArcGIS.⁶ The value called the condition number checks the level of local multicollinearity in the data. In case of strong evidence of multicollinearity, the results of the regression model will be unreliable, hence the variables being analyzed should not be larger than 30. The R Squared values indicate how well the model fits, and the values range from 0 to 1, indicating how well the model fits to the y values, thus a test of the fit of the model. Values closer to 0 has a poor fit. The output of the GWR analysis on ArcGIS produces a map of the Local R² values, to show where predictions were good and where they were not.

An alternative open-source software that can be used for the spatial autocorrelation analysis is R, although this has been developed recently⁷. Migration intensities, inequality data and County HDI can be exported from MS Excel to R using the.csv format. The county spatial maps are generated using the shapefiles, that can be read in QGIS or R package⁸. Linear regression analysis can be done using R package using the command *lm*. For the spatial autocorrelation analysis, which checks for clustering of migration intensities, the migration intensities and county shapefiles can be read in R. The ArcGIS analysis used the 'distance bands' to determine the nearest neighbor or contiguous county. In R package, one can use the 'contiguity neighborhood' as the measure of the connectedness of the counties, and calculate the mean values of the neighboring units. While ArcGIS runs the spatial autocorrelation using a weighted index, W, that captures the average weighted measure (spatial lag) of contiguous units, using R package, one can calculate the Moran's I using the *moran.test* function which gives you the Moran's I value and the p-values. Further information on spatial analysis using R can be obtained from several research works (see Baddeley *et al.*, 2016; Roger, *et. al*, 2013).

⁵The ArcGIS provides a tutorial on the interpretation of the spatial regression analysis, see <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-ols-regression-works.htm>

⁶<https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/interpreting-gwr-results.htm>

⁷<https://mgimond.github.io/Spatial/spatial-autocorrelation-in-r.html>

⁸<https://www.r-graph-gallery.com/168-load-a-shape-file-into-r.html>

Results

Correlation analysis was conducted between several inequality variables and migration intensity as the dependent variable using SPSS. The results are presented in [Table 1](#) and indicate that migration has a significant positive relationship ($p = 0.01$) with access to electricity, and with county development, County HDI ($p = 0.05$).

These findings corroborate other studies that show that migrants tend to move to areas with better development indicators, and the access to electricity is a good predictor of development in this case. Thus, regions with higher development indicators exhibit higher incidences of migration.

Ordinary Least Square (OLS) regression results

We conducted spatial analysis of our model, with migration intensity as the dependent variable, and the four explanatory variables namely access to electricity, access to water, County Gini and County HDI. The purpose of this analysis was to provide evidence of spatial association between the variables, and the consistency of such association within the counties. The spatial analysis results from the ordinary least square regression (OLS) reveal several observations. First, the nature and direction of the relationship between the variables are captured using scatterplots and histograms as shown in [Figure 1](#). The histograms show that the distribution of variables is not normal, with a left inclination, for the County Gini

Table 1. Pearson Correlation coefficient between migration effectiveness and indicators of development and inequality (N = 47).

	Migration effectiveness 2009	Proportion of households with electricity	Proportion of households without improved water	County Gini Coefficient	County HDI
Migration effectiveness 2009	1	.426**	-.336*	0.025	.321*
Proportion of households with electricity	.426**	1	-.674**	-0.11	.470**
Proportion of households without improved water	-.336*	-.674**	1	-0.122	-.372**
County Gini coefficient	0.025	-0.11	-0.122	1	-0.139
County HDI	.321*	.470**	-.372**	-0.139	1

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.05 level (2-tailed).

Figure 1. Scatterplot of variable distributions and relationships from OLS regression.

and access to power and a right inclination for the County HDI. The scatterplots show that the relationship between the variables and migration are non-linear. This corroborates literature findings implying that the relation between migration and inequality is curvilinear.

Table 2 presents diagnostic results of our OLS model. The Multiple R-squared value shows that our model explains up to 21 percent of the changes in migration intensity. Owing to the complexity of the relationship, we opt to use the Adjusted R-squared results, which show this efficacy level declines to 14 percent. The low explanatory level of our model implies that there may be other variables that could explain the variations in migration intensity, but these were excluded in our model. Despite this, our model is statistically significant, indicated by the Joint Wald statistic ($p < 0.05$). Now that we have our model and the efficacy confirmed, the OLS regression also tested if the observations of the relationship between the variables were consistent across the counties. The Koenker (BP) statistic checks for stationarity, and the results show a p value of 0.005, confirming that there is no heteroscedasticity in our model. Finally, the OLS results also confirm that the relationship between migration and inequality is non-linear, with the Jarque-Bera statistic value of 285 ($p < 0.001$) confirming this.

In **Table 3**, the OLS results show the explanatory powers of each of the variables in the regression. The results show that the County HDI and County Gini, have higher explanatory powers, although they influence migration intensity in opposite direction. The County HDI has a positive influence on migration intensity, while the County Gini has a negative influence on migration intensity. The remaining two variables, access to electricity and access to water, have lower explanatory powers in the equation. However, only the County Gini gives a statistically significant relationship when robust statistics is considered. A unit change in migration intensity results in a negative change of up to 567 units in County Gini, showing an inverse relationship between the two variables. This finding confirms earlier observations by **Kuznets (1971)** and other scholars of the inverse relationship between migration and income inequality.

Geographically Weighted Regression (GWR) results

The data was subjected to Geographically Weighted Regression (GWR) analysis and the attributes used in the analysis are shown in **Table 4**. With the GWR, the efficacy of our model improves and the model explains up to 20 percent of the relationship between migration and the inequality variables, as shown by the R^2 adjusted value. Just like the OLS model,

Table 2. Ordinary Least Square regression diagnostics for regression on migration effectiveness and the development and inequality variables.

Input features			Value
Number of observations			47
Multiple R-squared (d)			0.211
Adjusted R-squared (d)			0.136
Akaike's Information Criterion (AICc) [d]			640.358
Spatial derived statistics for goodness of fit statistics			
Joint F-Statistic (e)	2.81	Prob(>F), (4.42) degrees of freedom	0.038*
Joint Wald Statistic (e)	10.27	Prob(>chi-squared), (4) degrees of freedom	0.036*
Koenker (BP) Statistic (f)	15.05	Prob(>chi-squared), (4) degrees of freedom	0.005*
Jarque-Bera Statistic (g)	285.33	Prob(>chi-squared), (2) degrees of freedom	0.000*

P-value < 0.05, **p-value < 0.01.

Table 3. Summary of OLS results – model variables.

Variable	Coefficient (a)	Std Error	t-Statistic	Probability (b)	Robust_SE	Robust_t	Robust_Pr (b)	VIF (c)
Intercept	-59.24	296.47	-0.12	0.843	176.895	-0.335	0.739	-
HDI	386.16	438.58	0.88	0.384	334.688	1.154	0.255	1.323
GINI	-567.28	467.49	-1.21	0.232	204.306	-2.777	0.008*	1.128
ACC-Water	2.92	2.58	1.13	0.265	2.009	1.453	0.154	1.94
ACC-electricity	-8.74	2.82	-3.09	.004*	6.305	-1.386	0.173	2.083

Table 4. Gross Weighted Regression output and summary statistics.

OID	Variable name	Value	Definition
0	Bandwidth	990245.6	
1	Residual Squares	1508556	
2	Effective Number	6.21	
3	Sigma	192.3	
4	AICc	636.9	
5	R ²	0.29	
6	R ² Adjusted	0.20	
7	Dependent Field	0	Total Ages09
8	Explanatory Field	1	HDI
9	Explanatory Field	2	GINI
10	Explanatory Field	3	ACC_H20
11	Explanatory Field	4	ACC_ELEC

the efficacy confirms that several other critical variables that could explain migration and inequality are missing in our model.

The GWR results provide a cold-to-hot rendered map of the standard residuals of the regression analysis. This map, shown in [Figure 2](#), shows evidence of clustering of migration intensities in the country. For example, low migration intensities are found in Makueni, Machakos, Embu and Meru counties, respectively. Nairobi County has remarkably high migration intensity and is surrounded by regions with similarly high migration intensities, leading to a clustering of high-high migration. This may be because of the spillover effects of migration to Nairobi, hence migrants move to the next contiguous counties, as demonstrated by high intensities in Kajiado, Kiambu and Nakuru counties. In the western part of the country, there is a cluster of high migration in Vihiga County and similarly high migration in Migori County at the Kenya-Tanzania border. Comparatively, the coastal region shows evidence of low migration clustering in Mombasa and Kilifi counties, respectively.

Following the observed clustering of migration intensities in the country, a spatial autocorrelation of the residuals was applied to the data to confirm if the results are random, or clustered. The results of the Global Moran's I are presented in [Figure 3](#). Our Global Moran Index gives a value 0.105452, z-score is 3.0785, while p value is 0.002, implying that the data is spatially clustered and not randomly distributed. The positive value of Moran's I indicates that while the values are spatially clustered, positive values are clustered together and negative values are clustered together. This leads to the conclusion that migration intensity is spatially clustered, with neighboring regions recording similar values.

The results of the tests for identifying whether county migration is spatially clustered or not is presented in [Figure 4](#) which provides evidence of hotspots. The spatial clusters of migration intensities of similar nature confirm the existence of two key migration hotspots in the country. There is a high migration hotspot in the lake basin region (comprising of Kisumu, Vihiga and Nandi counties) and a low migration hotspot in the coastal region (comprising of Mombasa, Kilifi and Kwale counties). The Getis-Ord Statistic GI* show a positive value indicating that migration intensities are clustered in these regions, with positive values together and negative values together.

These finding confirms the importance of spatial factors in explaining demographic phenomena. In this case, the clustering of high migration intensities in the lake basin region confirms the previous observations of the region as a reservoir of migrants, having higher outflow of migrants largely due to scarcity of opportunities in the region. In a review of inequalities in Kenya, data showed the marginalization of the western part of the country owing to political reasons ([Ajulu, 2002](#)) resulting in lower human development outcomes ([Society for International Development, 2004, 2006](#)). For the coastal region, migration intensities are low with most movements towards Mombasa, the second largest capital city, and the counties that reported highest intensities are largely poorly resourced regions with high levels of poverty ([KNBS and SID, 2013](#)).

Discussion

This paper sought to test the relationship between migration and inequality in Kenya using several Subnational (county) variables including migration intensity, access to water, access to electricity and Human Development Index (County

Figure 2. Geographically Weighted Regression analysis.

HDI) and County Gini. We employed several tests including correlation and spatial regression techniques to make this determination. The former tests the general direction of the relationship between migration and different indicators of inequality, while the latter tests not just the association between the variables but checks for evidence of spatial relationships between the variables. The OLS results show that two measures of inequality, the County HDI and County Gini, have the highest explanatory powers for the changes in migration intensity. A statistically significant inverse relationship is also established for the County Gini and migration, such that a unit rise in migration intensity leads to a 567-unit decline in County Gini. This is a major contribution to the discourse on migration and inequality in Kenya.

The findings confirm that income is an important factor for migration, as indicated in earlier studies. These had shown that migrants tend to come from wealthier households, and that remittances sent by migrants increase income inequality between migrant and non-migrant households, as the migrant households receive remittances that increase their wealth status allowing them to invest in productive assets. While this may be true, our study did not test the family level

Figure 3. Spatial autocorrelation using Global Moran's I.

dynamics, but rather the meso level intensities. Our findings show that regions that report higher migration intensities, which implies that they have higher proportion of migrants in the total population, report lower levels of income inequality. This may imply that migration has an equalizing effect as a compensation mechanism for poor households as observed in several studies elsewhere (Kuznets, 1955; Arslan and Taylor, 2012; Mezger and Beauchemin, 2015).

Our spatial analysis shows that the migration intensities are not randomly spread in the country but cluster in specific geographic areas. The regions that have high migration intensity are clustered together, while those with low migration intensities are clustered together. As a result, Kenya has two key migration hotspots, a hotspot of high intensities at the lake basin region and a hotspot of low intensities at the coastal region. These results show that there are structural factors that may account for the variation of migration in the country and these underlying factors have an effect on the level of inequalities observed in the counties. We therefore conclude that spatial factors are important in understanding and interpreting the effects of migration on inequalities in Kenya.

The results of the GWR and OLS show that our model only partially explained the changes in the migration intensities in the country, thus the variables we chose to explain inequalities (County Gini, County HDI, access to water, access to electricity) may be insufficient in explaining the migration and inequality in the country. Our model only explained 20 percent of the variations, hence there may be other factors excluded as migration occurs as part of the wider social transformation system (Castels, 2010). Thus, for future research on migration and inequality, there may be need to consider the effect of other factors not included in this analysis.

The difference between our approach and that used by other scholars in previous studies was that we focused on the migration process measured by the level of intensity at the geographical level. The migration intensity measure, in our case, the Revised Weighted Net Migration Rate captures the overall effect of migration on population distribution, such that counties with high migration intensity are those with a higher proportion of internal migrants in the total population. The results of our analysis confirm that the counties with a higher proportion of migrants in their population tend to record significantly lower levels of income inequality. The 2009 migration intensities capture the changes that migration has on

Figure 4. Migration hotspots Kenya.

the receiving county populations, and the intensities were varied depending on the importance of migration to the county. For instance, our maps show a clustering of high migration intensities in Nairobi and surrounding counties, which is associated with higher development and economic opportunities, compared to the low intensities observed at the coast, a region associated with lower development.

Of the four variables we selected as explanatory variables, only income inequality measured by County Gini proved to be statistically significant in explaining the changes in migration intensities observed in the counties. Our findings corroborate previous observations by other scholars that when migration increases, there is first a rise in income inequality in the sending areas, which falls as more people migrate (see Kuznets, 1955). Some researchers suggest that the rise and eventual fall of income inequality with increased migration occurs because migrants send remittances to the sending communities, so as more migrants increase, there will be little or no income inequality gaps between migrant and non-migrant households (Stark and Taylor, 1991; Faini and Venturini, 1993; Vogler and Rotte, 2000).

While the migration intensities showed significant clustering, the inequality patterns measured by our four variables also depicted wide subnational variations. This may shed some light in the interpretation of our results. The regions with high income inequalities were mostly in the coastal region of Kenya especially in the counties of Tana River, Kwale and Kilifi. Coincidentally, these are also counties that experience high poverty levels ([Kenya National Bureau of Statistics and Society for International Development, 2013](#)). This may partly explain why the migration intensities at the coastal region are mainly low. The County Gini that is negatively associated with migration represents counties with higher income inequalities, some of which include the coastal region counties like Lamu, Kilifi, Kwale and Tana River, in Nyanza region including Siaya, Homa Bay, Kisii and Kisumu counties; in Busia County in Western Kenya and in Machakos County. The results imply that migration in Kenya is largely driven by the regional inequalities and corroborates previous scholarship ([Rempel, 1971](#); [Society for International Development, 2004, 2006](#); [Oucho, 2007](#); [KNBS and Society for International Development, 2013](#)) on the wide inequalities in the country.

Use of ArcGIS mapping enabled the visualization of migration intensities across the country, as well as mapping of the patterns of inequality. The results showed a north-south dichotomy in the patterns of inequality in the country, which may be traced to the country's colonial legacy. The patterns of migration in the country were spatially clustered and not randomized events.

Conclusion and recommendation

We conclude that our findings support previous observations on the inverse relationship between migration and inequality, albeit using different variables ([Bang, et al., 2016](#); [Wakajumma, 1986](#)). Using the ArcGIS and spatial analysis techniques, our results confirm that migration and inequality in Kenya have a spatial relationship, with migration patterns spatially distributed in response to the level of development in the country.

While we set out to confirm if we can test the effect of other dimensions of inequality and their relationship with migration, our results confirm that income inequality remains a robust measure of inequality and is negatively associated with migration in Kenya. The association between lower migration intensity and higher County Gini could reflect the differential impact of background factors, including structural factors, that affect who migrates and where they migrate to. The low explanatory power of the variables we chose in our study however show that additional factors need to be considered in the analysis of migration and inequality, especially those that influence the County Gini outcomes.

Our results confirm the importance of understanding not just the economic, but also the social and political contexts that affect mobility decisions. They point to the importance of a multi-layered analysis of migration and inequality in which structural and other factors are considered in the investigation. For example, we find that spatial factors remain important in explaining both migration and inequality in Kenya, as the patterns of migration and inequality show spatial variations.

Several recommendations can be made from this study. First, spatial analysis resulted in improvement of our understanding of the migration patterns in the country, hence it should be applied in future migration studies to improve the understanding of the migration process. The GWR and OLS analysis showed that the factors used in the model only explain 20 percent of the variations in the relationship between migration and inequality. Thus, we can explore migration beyond the income lens. For future analysis, we recommend that a mixed method approach is adopted in understanding how migration correlates with non-income inequalities.

A limitation of our study is the reliance on census data which only captures migration events without data on reasons for migration; therefore, it is difficult to determine individual level factors and confirm the migration history. Return migrants are also ignored in the analysis. While the data shows the overall impact of migration on the county, it was limited on the differential impacts within migrant households, migrant sending communities and the recipient communities.

Data availability

Underlying data

The internal migration dataset from the 2009 Kenya Population and Housing Census data is available from the Kenya National Bureau of Statistics (KNBS) and permission to obtain micro data can be obtained from their [website](#). Access to the data requires registration and is granted for those who wish to use the data for legitimate research purpose. Alternatively, the micro data is available from [Integrated Public Use Microdata Series \(IPUMS\)](#) following user request from their [registration page](#).

Inequality data was obtained from the Exploring Kenya Inequality National Report, published by the Kenya National Bureau of Standards (KNBS) and Society for International Development (2013). The report is available through this link: [Exploring Kenya Inequality National Report - Kenya National Bureau of Statistics \(knbs.or.ke\)](#).

County Human Development Index data was obtained from the 2009 Kenya National Human Development Report that can be accessed [here](#).

Access to electricity and access to water data was obtained from the Socio-Economic Atlas of Kenya (Wiesmann *et al.*, 2014) which contains analysis of the 2009 Kenya Population and Housing Census data. This can be accessed [here](#).

A guide for using R program for spatial autocorrelation can be found [here](#).

Acknowledgements

The late Professor John Oyaro Oucho contributed to the conceptualization of this study but did not see it to fruition.

References

- Adams RH Jr, Cuecuecha A, Page J: **The impact of remittances on poverty and inequality in Ghana**. *World Bank Policy Research Working Paper*. 2008; 4732.
[Publisher Full Text](#)
- Ajulu R: **Politicized ethnicity, competitive politics and conflict in Kenya: A historical perspective**. *African Studies*. 2002; **61**(2): 251–268.
[Publisher Full Text](#)
- Anselin L: **Spatial dependence and spatial structural instability in applied regression analysis**. *Journal of Regional science*. 1990; **30**(2): 185–207.
- Anselin L: **Local indicators of spatial association—LISA**. *Geographical Analysis*. 1995; **27**(2): 93–115.
[Publisher Full Text](#)
- Anselin L: **Local Spatial Autocorrelation**. 2017.
- Arslan A, Taylor JE: **Transforming rural economies: migration, income generation and inequality in rural Mexico**. *Journal of Development Studies*. 2012; **48**(8): 1156–1176.
[Publisher Full Text](#)
- Baddeley A, Rubak E, Turner R: *Spatial Point Patterns, Methodology and Applications with R*. Florida: CRC Press; 2016.
[Publisher Full Text](#)
- Bang JT, Mitra A, Wunna PV: **Do remittances improve income inequality? An instrumental variable quantile analysis of the Kenyan case**. *Economic Modelling*. 2016; **58**: 394–402.
[Publisher Full Text](#)
- Bell M, Blake M, Boyle P, *et al.*: **Cross-national comparison of internal migration: issues and measures**. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2002; **165**(3): 435–464.
[Publisher Full Text](#)
- Bivand RS, Gomez-Rubio V, Pebesma E: *Applied Spatial Data Analysis with R*. New York: Springer; 2013. Second
- Black R, Natali C, Skinner J: *Migration and Inequality*. Washington DC: World Bank; 2005.
- Castles S: **Understanding global migration: A social transformation perspective**. *Journal of Ethnic and Migration Studies*. 2010; **36**(10): 1565–1586.
[Publisher Full Text](#)
- De Haas H: **Remittances, migration, and social development. A conceptual review of the literature**. 2007.
- De Haas H: **Remittances and social development**. *Financing Social Policy: Mobilizing Resources for Social Development*. 2009; 293–318.
[Publisher Full Text](#)
- De Haas H: **Migration transitions: a theoretical and empirical inquiry into the developmental drivers of international migration**. 2010.
- De Haas H: *Migration theory: Quo vadis?*. International Migration Institute, University of Oxford; 2014.
- Ebeke CH, Le Goff M: **Why Migrants' Remittances Reduce Income Inequality in some Countries and not in Others?**. 2011.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Faini R, Venturini A: **Trade, aid and migrations: some basic policy issues**. *European economic review*. 1993; **37**(2-3): 435–442.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Fotheringham AS, Brunsdon C, Charlton M: *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons; 2003.
- Getis A: **Spatial autocorrelation**. *Handbook of applied spatial analysis*. Berlin, Heidelberg: Springer; 2010; 255–278.
- Getis A, Ord JK: **The analysis of spatial association by use of distance statistics**. *Geographical Analysis*. 1992; **24**: 189–206.
[Publisher Full Text](#)
- Hepple LW: **Exact testing for spatial autocorrelation among regression residuals**. *Environment and Planning A*. 1998; **30**(1): 85–108.
[Publisher Full Text](#)
- Hoddinott J: **A model of migration and remittances applied to Western Kenya**. *Oxford economic papers*. 1994; **46**: 459–476.
[Publisher Full Text](#)
- Kenya National Bureau of Statistics (KNBS) and Society for International Development (SID): *Exploring Kenya's Inequality: Pulling Apart or Pooling Together?*. Nairobi: Kenya National Bureau of Statistics; 2013.
- King R: **Theories and typologies of migration: an overview and a primer**. 2012.
- Klugman J: **Human development report 2009. Overcoming barriers: Human mobility and development**. 2009.
- Knowles JC, Anker R: **An analysis of income transfers in a developing country: The case of Kenya**. *Journal of Development Economics*. 1981; **8**(2): 205–226.
[Publisher Full Text](#)
- Kuznets S: *Notes on Stage of Economic Growth as a System Determinant*. University of California Press; 1971; vol. **8**: 243–268.
- Liu S, Hu Z, Deng Y, *et al.*: **The regional types of China's floating population: Identification methods and spatial patterns**. *Journal of Geographical Sciences*. 2011; **21**(1): 35–48.
[Publisher Full Text](#)
- Melamed C, Samman E: *Equity, inequality, and human development in a post-2015 framework*. UNDP, Human Development Report Office; 2013.
- Mezger Kveder C, Beauchemin C: **The role of international migration experience for investment at home: Direct, indirect, and equalising effects in Senegal**. *Population, Space and Place*. 2015; **21**(6): 535–552.
[Publisher Full Text](#)
- Muyonga M, Odipo G, Agwanda AO: **Interlinkages between Migration and Inequality in Africa: Review of Contemporary Studies**. *AHMR*. 2020; 6.
- Ord JK, Getis A: **Local spatial autocorrelation statistics: distributional issues and an application**. *Geographical analysis*. 1995; **27**(4): 286–306.
[Publisher Full Text](#)
- Oucho JO: **Migration and regional development in Kenya**. *Development*. 2007; **50**(4): 88–93.
[Publisher Full Text](#)
- Oyvat C, Wa Githinji M: **Migration in Kenya: Beyond Harris-Todaro**. *Department of Economics Working Paper Series*. 2017; 218.
- Pearson K: **Determination of the coefficient of correlation**. *Science*. 1909; **30**(757): 23–25.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Plaza S, Navarrete M, Ratha D: *Migration and remittances household surveys in sub-Saharan Africa: methodological aspects and main findings*. Washington, DC: World Bank; 2011.
- Rees P, Bell M, Duke-Williams O, *et al.*: **Problems and solutions in the measurement of migration intensities: Australia and Britain compared**. *Population Studies*. 2000; **54**(2): 207–222.
[Publisher Full Text](#)
- Rempel H: *Labor migration into urban centers and urban unemployment in Kenya*. PhD Dissertation. University of Nairobi, Kenya; 1971.

Republic of Kenya: *2009 Kenya Population and Housing Census, Analytical Report Volume VII on Migration*. Nairobi: Kenya National Bureau of Statistics; 2012.

Shi L, Chen W, Xu J, et al.: **Trends and Characteristics of Inter-Provincial Migrants in Mainland China and Its Relation with Economic Factors: A Panel Data Analysis from 2011 to 2016**. *Sustainability*. 2020; **12**(2): 610.

Shrestha N: **Detecting multicollinearity in regression analysis**. *American Journal of Applied Mathematics and Statistics*. 2020; **8**(2) (2020): 39–42.

[Publisher Full Text](#)

Society for International Development (SID): *Pulling Apart: Facts and Figures on Inequality in Kenya*. Nairobi: SID Eastern Africa Regional Office; 2004.

Society for International Development (SID): *Inequality in Kenya*. Nairobi: SID Eastern Africa Regional Office; 2006.

Stark O, Taylor JE: *Relative deprivation and migration: theory, evidence, and policy implications*. World Bank Publications; 1991; vol. **656**.

Tiefelsdorf M, Boots B: **The exact distribution of Moran's I**. *Environment and Planning A*. 1995; **27**(6): 985–999.

[Publisher Full Text](#)

Tiefelsdorf M, Boots B: **A note on the extremities of local Moran's Is and their impact on global Moran's I**. *Geographical Analysis*. 1997; **29**(3): 248–257.

[Publisher Full Text](#)

Tobler WR: **A computer movie simulating urban growth in the Detroit region**. *Economic geography*. 1970; **46**(sup1): 234–240.

[Publisher Full Text](#)

United Nations: *Human Development Report 2005*. 2005.

United Nations: **Human Development Report 2009. Overcoming Barriers: Human Mobility and Development**. 2009.

United Nations Development Program (UNDP) Kenya: *Kenya National Human Development Report 2009: youth and human development: tapping the untapped resource*. Nairobi: UNDP Kenya; 2010.

United Nations: **Human Development Report 2013: Humanity divided: Confronting inequality in developing countries**. 2013.

Van Imhoff E, Keilman N: *LIPRO 2.0: An application of a dynamic demographic projection model to household structure in the Netherlands*. (Vol. 23). Amsterdam: Swets & Zeitlinger; 1991.

[Publisher Full Text](#)

Vogler M, Rotte R: **The effects of development on migration: Theoretical issues and new empirical evidence**. *Journal of Population Economics*. 2000; **13**(3): 485–508.

[Publisher Full Text](#)

Wakajumma JO: *Intercensal net migration in Kenya district level analysis*. (Doctoral dissertation). University of Nairobi; 1986.

Wartenberg D: **Multivariate spatial correlation: a method for exploratory geographical analysis**. *Geographical Analysis*. 1985; **17**(4): 263–283.

[Publisher Full Text](#)

Wiesmann UM, Kiteme B, Mwangi Z: *Socio-economic atlas of Kenya: Depicting the national population census by county and sub-location*. Kenya National Bureau of Statistics, Centre for Training and Integrated Research in ASAL Development, Centre for Development and Environment; 2014.

Young DS: *Handbook of regression methods*. CRC Press; 2018.

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 13 December 2021

<https://doi.org/10.5256/f1000research.77770.r101215>

© 2021 Nzabona A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abel Nzabona

Department of Population Studies, Makerere University, Kampala, Uganda

Introduction:

- Authors have a detailed introduction that delves into the subject matter. The section would be enriched if key concepts (such as intensity of migration, inequality e.t.c.) were operationally defined.

Data and methods:

- Authors are using data collected some 12 years ago in the 2009 national census. The 2019 census is obviously more recent and would be expected to be the ideal choice. If there are reasons for not using it, let these be brought out to justify the use of the older dataset.
- It is helpful to make solid and clear justification for other choices. These include selecting **water** and **electricity** as explanatory variables. Why not other socioeconomic variables but these two? Similarly, opting for lake basin and coastal areas requires explanation.
- At the end of the section labelled 'analytical methods', there is a long paragraph about 'an alternative open-source software that can be used for the spatial autocorrelation analysis...' and this is in reference to 'R'. What is the point in disclosing possibilities in a section like this? This section is expected to touch on what 'was done and how' rather than what 'can be done and with what'.

Results:

- Use of map displays is commendable as the results are presented in an interesting and elegant manner. Tables are used mainly for methodological results. It would be appreciated if they were similarly used for substantive results on migration-inequality nexus.

Recommendations:

- Relevant methodological recommendations are made. However, less is done regarding policy implications and recommendations that emerge from substantive results. Authors need to re-visit this area.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Demographics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research