# Socio-economic Factors that Determine Employment in Tanzania: Spatial Analysis

Godwin Aloyce Myovella

Department of Economics and Statistics, University of Dodoma, Tanzania

E-mail: myovellagd@gmail.com

## Abstract

This paper investigates the socio-economic determinants of Employment in Tanzanian regions. It used the cross-sectional data of 2014 on several variables as this was found to be the most recent dataset in the Tanzania National Bureau of Statistics. Despite the impressive growth rates of 7 percent observed in Tanzania poverty hasn't been reduced as expected, and employment has been static. The paper applies spatial techniques including spatial lag (SAR), spatial error (SEM) and spatial Durbin models to discover the spatial patterns of employment. SAR model was selected as all the variables, GDP per capita, the number of schools, and health services measured by a number of dispensaries were significant. Also, the direct and indirect effects were conducted and GDP per capita were significant and conforms to the theory. The study concludes that there was some regional dependency on some variables hence spatial approach is appropriate as it avoids biased, inefficient and inconsistent results.

**Keywords:** Socio-Economic, Employment, Tanzania, Spatial model

## 1. Introduction

For about a decade there has been an impressive economic growth of about 7 percent in Tanzania, however, it has been cited that this growth hasn't been able to reduce poverty amongst the population as it has been expected. According to Tanzania Human Development Report, both transformation of economic growth and essential change in the economic structure, as well as the creation of employment, is necessary URT, (2014). Further, Pavlyuk, (2011) asserts that, to realize economic wellbeing and growth potential employment rates are extensively known indicators. There has been an increase of the population of Tanzania by about 1.2 million people annually and the labor force has been increasing at an annual rate of 2.3 percent but for a long time employment has been said to be static.

One of the important problems of development in Tanzanian is the inter-regions development

disparities which are widely mentioned. With this nature, it can be argued that there is a spatial dependency in the regional employment as well. So far we have not come across any study conducted in Tanzania using regional data and spatial econometric techniques. The objective is to study the association between regional employment and socioeconomic variables in Tanzania using the spatial analysis. I consider the spatial approach to be appropriate because I believe regional labor market has strong spatial relationships with neighboring regions.

## 2. Literature Review

Many authors have asserted that unemployment is a major problem in almost all countries of the World (Kantar and Gunay, 2016). However, mostly unemployment follows after the recession's episodes but it has become persistent sometimes in some countries, see (ILO, 2010a; Arpaia and Curci, 2010; Choudhry et al., 2012; O'Higgins, 2012). Europe, for instance, has been suffering a prolonged slump with high levels of unemployment posing a threat to the political and social stability of the European Union (Marques& Lima, 2015). It has become even worse in the recent past for example in the European countries after the global financial and economic crisis.

The negative effects of high unemployment need not to be overemphasized. Olga, et al, (2009) stress that unemployment is mentioned to be one of the major failures of market economies, they further argued that not only is it a cost for society but the cost is heavier for specific categories of people, since the unemployment rate is widely differentiated by gender (in many countries it is higher for women), by age, by country or region and also it changes over time.

Despite the fact that many countries have a concern on high unemployment rates as an aggregate variable, on the other hand, there is a problem of employment disparity among regions within a country which is even less studied (Filiztekin, 2009) see also (Lopez-Bazo, etal, 2002). Consequently, the analysis of regional unemployment differences has attracted increasing interest in the economic literature (Maria & Miranda, 2007).

This particular study focuses on regional disparities in employment among regions in Tanzania as motivated by the study of Filiztekin and other authors who did similar works in regional employment. In a nutshell, a brief review of the literature is made on the subject, of interest is that most of the studies which are reviewed are those that used spatial techniques in their analysis as a similar approach will be used in this study.

Kantar and Gunay (2016) conducted a study aimed at investigating the spatial patterns of regional unemployment rates in Turkey for various years 2004, 2011, 2012 and 2013.They focused on a provincial level and used local and global spatial autocorrelation statistics, also mapping analysis methods. In their study, it was found that there is a significant neighboring effect among unemployment rates at the provincial level in Turkey over time. In other words, neighboring provinces in Turkey showed similar values in terms of unemployment rates, suggesting some spatial dependency at the provincial level.

Different from the approach by Kantar and Gunay in Turkey, Filiztekin, (2009) conducted a

study before them. Spatial and nonparametric techniques were used and revealed the wide regional unemployment disparities in Turkey. The data used was from 1980 to 2000 and it showed that the provincial unemployment rates were quite persistent and the gap across different regions widens even further with spatial clusters emerging across the country. The evidence indicates that human capital and demand deficiency are the sources of observed disparity across provinces.

Demidova and Signorelli, (2012) conducted a study is to identify the common and different determinants of youth unemployment in Eastern and Western regions of Russia, especially searching for the existence of spatial effects. For them, they were much interested in the youth unemployment because they argue that youth unemployment in many countries is usually higher than the total unemployment. They, therefore, used the data for 75 Russian regions from 2000 to 2009 to study the spatial effect among other research questions they had and they used the panel model. Finally, their results showed the significance of spatial lag coefficients.

In spatial research on regional employment rates in Latvia, Pavlyuk (2011), used a set of model explanatory variables which included all factors, related to employment in a region. The focus was much on the analysis of local employment in the districts of Latvia. They utilized a modern approach to discover spatial patterns of employment rates. A spatial econometric model is presented in the research, and its parameters are estimated on the basis of Latvian regional statistical data. They discovered significant relationships between employment and regional economic indicators and also a strong spatial relationship between employment rates in the districts of Latvia. They considered a set of region-specific factors and related indicators as employment determinants: regional GDP per capita, a number of beds in hospitals per capita, a number of crimes per capita, flat prices, a number of schools and preschools per capita and average wages.

In all these studies and many others, there have been widely used a number of explanatory variables which have influence and have been used to study the regional unemployment. Among the factors on the list are; the natural change in the labor force, the participation rate, net in-migration and commuting, wages, employment growth, the industrial mix, the educational attainment of the population, market potential, and other characteristics of the labor market such as the degree of unionization. This study used the above-mentioned papers to select the variables used especially those similarly used in the Pavlyuk. Some of them have not been used due to data unavailability in Tanzanian context.

Anselin (2006) asserts that generally in recent years the interest in spatial analysis and spatial econometric analysis, in particular, has seen an almost exponential growth, especially in the social sciences.

## 3. Model Specification

According to the main goal of the study the dependent variable ($Y_j$) will be regional employment and the independent variables will be, ($X_1$) regional GDP per capita (at current market prices in Tanzanian shillings), ($X_2$) regional total crimes (which includes offenses to

property offence against person and offence against public tranquility, ($X_3$) number of schools, ($X_4$) health services.

The model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

In a compact form, we can write our OLS model as

$$y = X\beta + \varepsilon$$

Where;

$Y$ is a vector of a dependent variable;

$X$ is a matrix of explanatory variables;

$\beta$ is a vector of unknown coefficients (model parameters);

$\varepsilon$ is a vector of i.i.d. error terms

We expect the signs to be positive with the economic development of a region (GDP per capita) on employment. Also, we expect a positive correlation with social conditions (number of schools, number of hospitals) and a negative correlation with crimes.

The classical regression model is the simplest model and is the most frequently used method for investigation of the stochastic relationship between a dependent variable and a set of the determinant. However if we expect the presence of the spatial effects, classical OLS estimates will be biased (so estimates, directions of influence, estimates significance, and other model characteristics will be wrong). Therefore we used three models in spatial effects such as spatial lag, spatial error, and spatial Durbin

**Spatial lag and spatial error models**

There are two main components of spatial dependency. Firstly, spatial dependency is presented in form of influence of economic characteristics in neighbor regions on employment in a given region. This form of dependency is usually called spatial lags. The direction of spatial lags influence on employment rate can be both negative and positive, so we don't have strong a priori expectation here.

We can express a spatial lag model more formally as

$$y = \rho W y + X\beta + \varepsilon$$

Where y is a vector of a dependent variable; X is a matrix of explanatory variables; W is a spatial weight matrix; $\rho$ is unknown coefficients reflecting spatial influence; β is a coefficient of the independent variables; ε is a vector of i.i.d. error terms and Wy is correlated with the disturbances.

Another form of spatial dependency is a correlation of error terms in neighbor regions (called spatial errors). Consequently, OLS remains unbiased, but it is no longer efficient and the

classical estimators for standard errors will be biased. The spatial error can be expressed as

$$y = X\beta + \varepsilon(I - \lambda W)$$

Where $\lambda$ is autocorrelation parameter in residuals

**Spatial Durbin model**

This is a special case of Spatial Auto regressive model (SAR) Spatial Durbin Model which add lag effect of the independent variables so that the model is as following

$$y = \rho W y + X\beta + W X\theta + \varepsilon$$

## 4. Data Sources and Description of Area of Study

This study used secondary data collected from the Tanzania National Bureau of Statistics database. Regional employment data were obtained from the database and the regional employment rate was calculated by the author by taking the ratio of the employed to the total population of the region. All other data or indicators used were obtained directly from the database. The study period was 2014 the reason is the availability of most data in this year, therefore, the author considers it to be recent according to the updates of data in the Nation's database.

The Area of the study was Tanzania mainland and it covered 21 regions 5 regions will not be covered under study because they are newly established administrative regions and there is no enough data of the selected variables under study.

Table 1. Descriptive Statistics of the Data

|  | GDP PER CAPITA | CRIMES | EMPLOYEES | DISPENSARY | CHOOLS |
|---|---|---|---|---|---|
| **Mean** | 1672061. | 3012.286 | 98803.76 | 256.4762 | 206.7143 |
| **Median** | 1656913. | 2191.000 | 48183.00 | 253.0000 | 191.0000 |
| **Maximum** | 2797694. | 18516.00 | 720480.0 | 389.0000 | 344.0000 |
| **Minimum** | 1023631. | 1138.000 | 18219.00 | 157.0000 | 91.00000 |
| **Std. Dev.** | 482213.8 | 3640.877 | 152902.3 | 68.53438 | 70.86829 |
| **Skewness** | 0.503236 | 3.920366 | 3.455318 | 0.346570 | 0.552043 |
| **Kurtosis** | 2.719062 | 17.27813 | 14.43338 | 1.997867 | 2.382174 |
| **Jarque-Bera** | 0.955422 | 232.1744 | 156.1692 | 1.299126 | 1.400624 |
| **Probability** | 0.620202 | 0.000000 | 0.000000 | 0.522274 | 0.496430 |
| **Sum** | 35113277 | 63258.00 | 2074879. | 5386.000 | 4341.000 |
| **Sum Sq. Dev.** | 4.65E+12 | 2.65E+08 | 4.68E+11 | 93939.24 | 100446.3 |
| **Observations** | 21 | 21 | 21 | 21 | 21 |

## 5. Model Estimation and Hypothesis Testing

### 5.1 Spatial Weight Matrix (W)

Although LeSage showed that there is a small amount of discrepancy in spatial contiguity matrices constructed by MATLAB® versus contiguity matrices constructed by hand, in this

analysis, the spatial element was constructed through the use of LeSage's MATLAB® spatial econometrics toolbox and latitude/ longitude coordinates.

*5.2 Moran's I*

The first step is to discover possible spatial effects in our sample. The most commonly used specification test for spatial autocorrelation is derived from a statistic developed by Moran (1948). Moran's I coefficient allows discovering spatial autocorrelation of model residuals, but cannot distinguish between spatial lags and spatial errors. In this test, we set Moran's I value to be zero as a null hypothesis. Then, there is no spatial effect if Moran's I value is significantly zero. In matrix notation, Moran's I statistic is

$$I = \hat{\varepsilon}'W\varepsilon/\hat{\varepsilon}\hat{\varepsilon}' \sim i.i.d\ normal$$

In this test, a null hypothesis is, H0: there is no spatial autocorrelation. If it cannot be rejected, there is no need for a spatial econometric model.

Table 2. Moran I test for spatial correlation in residuals

| Moran I | 0.9968 |
|---|---|
| Moran I statistic | 12.4478 |
| Marginal probability | 0.0000 |

Source: Authors Calculation

Moran's value is significantly different from zero, a null hypothesis is thus rejected. It shows that regional employment rate is positively spatially auto-correlated during 2014. However, from this test only, we could not say anything about the model; we knew just there is the spatial effect.

**Diagnostic test for spatial dependence**

We have now to determine whether there is spatial lag or spatial error. In that aim, we performed Lagrange Multiplier (LM) of error and lag in Matlab.

Table 3. LM test for spatial correlation on SAR Model

| LM value | 3.8647 |
|---|---|
| Marginal Probability | 0.0493 |
| Chi (1) .01 value | 6.6350 |

Source: Authors Calculation

Table 4. LM test on SEM model

| LM value | 1.8071 |
|---|---|
| Marginal Probability | 0.1788 |
| Chi (1) .01 value | 17.6110 |

Source: Authors Calculation

## 6. Model Estimation

The spatial econometric literature provides different models for data with spatial autocorrelation; we have three models such as SAR, SEM, and SDM. We have also run the ordinary least square test in our analysis, and the table below summarizes the test result.

Table 5. Estimation Test Result

| Parameter | OLS | SAR | SEM | SDM |
|---|---|---|---|---|
| Constant | -26.35** (-5107) | -23.438**(-5.658) | -23.86**(-5.452) | -61.19** (-3.80) |
| GDP/c | 1.281**(3.55) | -1.130**(3.985) | 1.106**(3.604) | 1.376** (4.589) |
| Crime | 0.09(0.413) | -0.05(0.295) | 0.046(0.249) | 0.131 (0.749) |
| Health (dsp) | 1.23(1.710) | -0.24**(-0.567) | 1.0035*(1.727) | 1.33** (2.21) |
| Edu (sch) | -0.431(-0.787) | 1.148**(2.058) | -0.1916(-0.4168) | -0.542 (-1.45) |
| Rho/Lamda | - | 0.36**(2.163) | 0.36 (1.372) | -0448 (-1.366) |
| W*GDP | - | - | - | 1.78** (2.144) |
| W*Crime | - | - | - | 0.55 (1.41) |
| W*Edu (sch) | - | - | - | 0.20 (0.187) |
| W*Health (dsp) | - | - | - | 0.125 (0.082) |
| Log-likelihood | - | 0.328 | -1.61 | 3.096 |

Source: Authors calculation ( ): t-statistics

Table 7. Direct and indirect Effect

| Variable | Direct effect | t-probability | Indirect effect | t-probability |
|---|---|---|---|---|
| GDP | 1.183 | 0.00072** | 0.751 | 0.256 |
| Crime | 0.049 | 0.784 | 0.02 | 0.843 |
| Edu (sch) | -0.24 | 0.578 | -0.138 | 0.702 |
| Health (disp) | 1.196 | 0.052* | 0.971 | 0.366 |

Source: Authors calculation

## 7. Interpretation of the Results

In the first step of the analysis, Moran's test rejected the null hypothesis (no spatial effect) in the results indicated above. This means the presence of the spatial effect in our sample; therefore, we apply spatial econometrics throughout the analysis.

The next question was raised on whether to use spatial lag or spatial error. In the diagnostic section, we could answer this question. It was only significant for SAR model after using langrage multiplier test in the analysis (see in table 3). Both of the models were estimated however and from the parameters, the lambda in error model was not significant therefore it was concluded that the best model is SAR. We performed the test estimation for both model and spatial Durbin model as well using Matlab.

Firstly, we conducted the ordinary least square and the results show all the variables were insignificant except for the GDP per capita. The results which were not expected and they are

different from other similar studies which used similar explanatory variables in other countries. The crime was insignificant and it had no negative sign as it was anticipated from the beginning. However, we continued with the spatial model estimations. In SAR model all the explanatory variables were significant except the crime, however, this time it had a negative sign. GPD per capital, the number of school and the health services here taken as the total number of dispensaries in each region were significant as shown in the table above.

From the LM test of maximum likelihood, we conducted the test of the SAR and SEM model and it was significant for the former while the latter was insignificant that led me to choose SAR model as the best model (appropriate) of this study. It is from that model I conducted the direct and indirect effect and from the direct-indirect effect table GDP per capital and health services were significant and conform to the theory. We can say that a one percent increase in the GDP per capital in the region has an effect on the employment of 1.83 percent in that region. The reason is that it is the employees who contribute to the GDP. Same applied to the health services a one percent increase in the health services leads to an increase of 1.196 percent in employment in the particular region. However, due to insignificant results to the indirect effect, we may argue that an increase in the GDP per capital in a neighboring region does not lead to an increase in employment in the region. Then we express the model as

$$Y = 0.36Wy - 1.13X_1 - 0.05X_2 + 1.14X_3 - 0.24X_4 + \hat{\varepsilon}$$

Where

$Y_t$ : Employment and the independent variables will be,

$X_1$ : Regional GDP per capita (at current market prices in Tanzanian shillings),

$X_2$ : Regional total crimes (which includes offenses to property offense against person and offense against public tranquility.

$X_3$ : Number of schools, $X_4$ : Health services (dispensaries)

## 8. Conclusion

The main objective of this study was to investigate the socio-economic determinants of employment in Tanzanian regions. The motivation behind was that we have not come across similar studies with the modern spatial approach which has been done in Tanzania only in some developed countries. We discovered that there was some regional dependency in some variables and therefore we can conclude that if a similar study is to be conducted without considering some Spatial approaches the results may be biased, inefficient and inconsistent.

Our results show that the Spatial Lag (SAR) model fitted our data well however the unexpected results such as the signs on crimes, Health and GDP in the SAR model may have arisen due to using cross-section data. It would be better to use panel or pooled OLS model to obtain the better results. The unavailability of data in The National database of Tanzania for some years is what led us to the selection of the data in that particular year.

# Reference

Anselin, L. (January 01, 2006). How (not) to lie with spatial statistics. *American Journal of Preventive Medicine, 30*(2), 3-6. https://doi.org/10.1016/j.amepre.2005.09.015

Anselin L. (2006). Spatial Analysis Laboratory, Department of Geography and National Centre for Supercomputing Applications, University of Illinois, Urbana, IL 61810.

Arpaia, A., Curci, N., & European Commission. (2010). EU labour market behaviour during the Great Recession. Brussels: European Commission, Directorate-General for Economic and Financial Affairs.

Choudhry, M. T., Marelli, E., & Signorelli, M. (January 01, 2012). Youth unemployment rate and impact of financial crises. *International Journal of Manpower, 33*(1), 76-95. https://doi.org/10.1108/01437721211212538

Demidova, O., & Signorelli, M. (June 01, 2012). Determinants of youth unemployment in Russian regions. *Post-communist Economies, 24*(2), 191-217. https://doi.org/10.1080/14631377.2012.675155

Filiztekin, A. (January 01, 2009). Regional unemployment in Turkey. *Papers in Regional Science, 88*(4), 863-878. https://doi.org/10.1111/j.1435-5957.2009.00237.x

ILO (2010a). Global Employment Trends. Geneva.

Kantar, Y. M., & Aktaş, S. G. (June 22, 2016). Spatial Correlation Analysis of Unemployment Rates in Turkey. *Journal of Eastern Europe Research in Business and Economics*, 1-9.

Lopez-Bazo, E., del Barrio, Castro, T., & Artis, M. (2002). The regional distribution of Spanish unemployment: A spatial analysis, *Papers inRegional Science, 81*(3), 365-389.

Maria, F., Miranda, C., & Nijkamp, P. (December 01, 2007). Geographical Distribution of Unemployment: An Analysis of Provincial Differences in Italy. *Growth and Change, 38*(4), 649-670. https://doi.org/10.1111/j.1468-2257.2007.00391.x

Marques, M., & Lima, G, (2015). Testing for Spatial Dependence in the Regional Unemployment in Europe.

O'Higgins, N., & International Social Security Association. (January 01, 1997). The challenge of youth unemployment [in: Unemployment and social security: special issue]. *International Social Security Review, 50*(4), 63-93. https://doi.org/10.1111/j.1468-246X.1997.tb01084.x

Olga, D., Marelli, E., & Signorelli, M. (2009). The Spatial Effects of the Youth Unemployment Rate for Russia and Italy, National Research University Higher School of Economic, 2-12

URT. (2014). Human Development Report: Sustaining Human Progress, Reducing Vurnerabilities and building Resilience. New York: UNDP

**Copyright Disclaimer**