A spatially varying relationship between the proportion of foreign citizens and income at local authorities in Greece

Stamatis Kalogirou[1]
[1] *Department of Geography, Harokopio University, El. Venizelou 70, Kallithea 17676, Athens,* skalo@hua.gr

Abstract
The aim of this paper is to spatially examine the connection between the proportion of foreign citizens and household income at local authority level in Greece. Since the majority of foreign citizens in Greece are economic immigrants, it is interesting to understand if their presence at municipality level is associated with high household income or poverty. One could expect either a positive or a negative association for difference reasons. The traditional correlation coefficient suggests a positive and significant relationship. However, when this is examined locally, using a geographically weighted Pearson correlation coefficient, the relationship is spatially non-stationary. There are touristic areas where the relationship is much stronger. The latter non-stationarity is sensitive to the size of the spatial kernel; the smaller the kernel size is, the wider the range of the local coefficients is.

Keywords
Migration, Income, Spatial Analysis, Local correlation, lctools

1. Introduction and Background

Historically, since the 1920s, Greece has been a country producing emigrants, thus the large Greek Diaspora across the world. However, since the late 1980s, and especially in recent years (post soviet era), Greece became a destination for immigrants from East Europe, Asia and Sub-Saharan Africa resulting in a growing foreign population in absolute and proportional terms. There have been several issues related to immigrants in Greece based on census and survey data. For example, Rovolis and Tragaki (2006; 2008) examine the geographical distribution of immigrants and their socio-demographic and ethnic characteristics.

This paper is concerned with a potential relationship between the proportion of the non-native Greek population and the mean recorded household income in Local Authorities (LAs) in Greece. Answering this question will help understand better the residential choice by immigrants in relation to the local economic conditions of the general population. It can equally be expected that immigrants could live in prosperous LAs which potentially offer more employment opportunities (a positive association between foreigners and income) or deprived LAs with more affordable housing (a negative association between foreigners and income). However, after considering that in Athens, where the majority of foreigners reside a vertical segregation is evident (Maloutas and Karadimitriou 2001), one might expect that a positive association between the proportion of foreign citizens and income is more plausible.

The above research question is not to be confused with the economic conditions of foreigners in relation to native Greeks. Evidence can be provided that on average, foreigners' income is lower than natives' income. For example, based on the Statistics on Income and Living Conditions Survey (EU-SILC) in households of year 2003 for Greece, the mean equivalized disposable income of native Greeks was 9779.62 Euros per person while that of non-natives was 8297.83 Euros.

A simple way to answer the above question is by calculating a correlation coefficient between the two variables. However, such a measure may be misleading when spatial data are concerned. For example, income data in Greece exhibit positive spatial autocorrelation resulting in spatial patterns of high income in urban areas and low income in rural and remote areas (Kalogirou 2010). It makes sense to expect that a variable association between income and foreigners is more plausible than a single association that a traditional coefficient would suggest.

In this paper, data from the 2001 Population Census and the 2002 tax forms registration aggregated at the local authority level in Greece are analysed in order to answer the above question. The analysis involves calculating local and geographically weighted correlation coefficients in order to assess the spatial variation of the relationship in question. The following sections present further information on the nature and source of the data. The methodology on local correlation coefficients and the results of the analysis are then presented in detail. The conclusion of this research and plans for future analysis follow.

2. Materials and Methods

The mean recorded household income refers to the average household income earned during the calendar year 2001 recorded in the year 2002 for each postal code in Greece aggregated to the Local Authority level. The term “recorded” in this case refers to the household income declared in tax forms inflated according to the wealth of the households, such as properties and equities (Kalogirou and Hatzichristos 2007). The source of income data is the General Secretariat for Information Systems (GSIS) of the Ministry of Finance of the Hellenic Republic. The data at the postal code geography was aggregated to the 1033 Local Authority geography in Greece (Kalogirou 2010) that the 2001 Census referred to.

The proportion of people that do not have the Greek citizenship, also referred to as foreigners or immigrants interchangeably in this paper, include legal immigrants and their descendants who may have been born in Greece but do not have the Greek citizenship. The data source for citizenship is the 2001 Census for Population in Greece.

It is important to recognise that in terms of the data on foreigners, illegal immigrants are unaccounted for and recorded income is usually lower than the truly generated income due to the large extend of the black economy in Greece (Kalogirou, 2010).

A spatial database including these two variables is publicly available as GR.Municipalities in the R package ***lctools*** (Kalogirou, 2015). Maps of the proportion foreigners and mean recorded household income are presented in Figures 1 and 2, respectively.

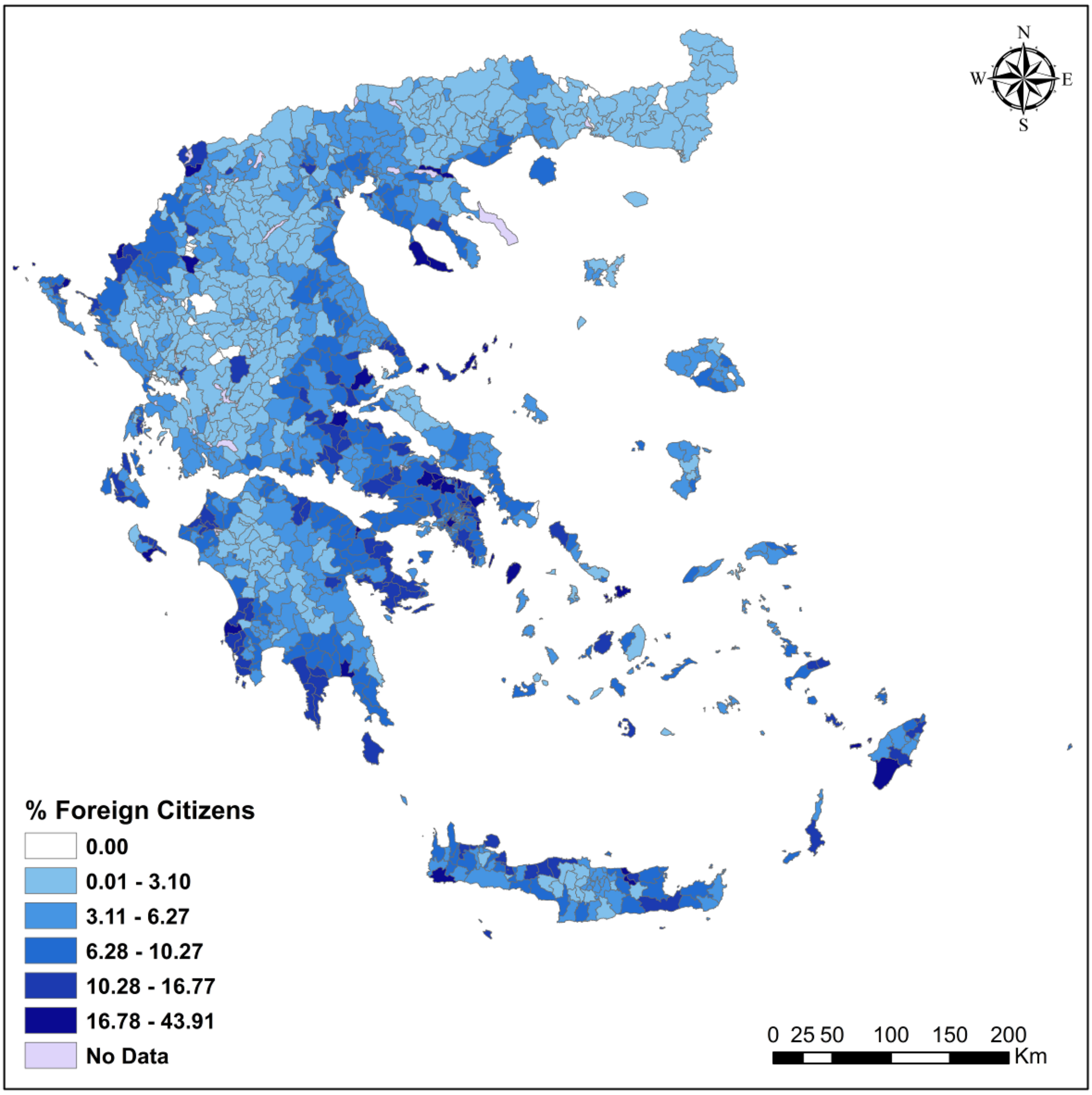

Figure 1. Proportion foreigners for Local Authorities in Greece

The most common statistic to measure the association between two variables, that does not necessarily imply a causational relationship, is a standard correlation coefficient such as the Pearson Correlation Coefficient (PCC). However, such a statistics is "global" and thus likely to hide significant spatial variation of the relationship between two variables. The latter could be identified if a local version of the correlation statistics was calculated. Such local versions could be the Local Pearson Correlation Coefficient (LPCC) and the Geographically Weighted Pearson Correlation Coefficient (GWPCC) a short description of which follows. These local statistics have already been proposed to assist, among others, in the statistical inference of regression analysis (Kalogirou 2012). Software to calculate these statistics as well as significance tests is available as a free standalone windows application (LC-Tools) and R package (lctools), in http://gisc.gr/software/lc-tools/ and http://cran.r-project.org/web/packages/lctools/index.html, respectively.

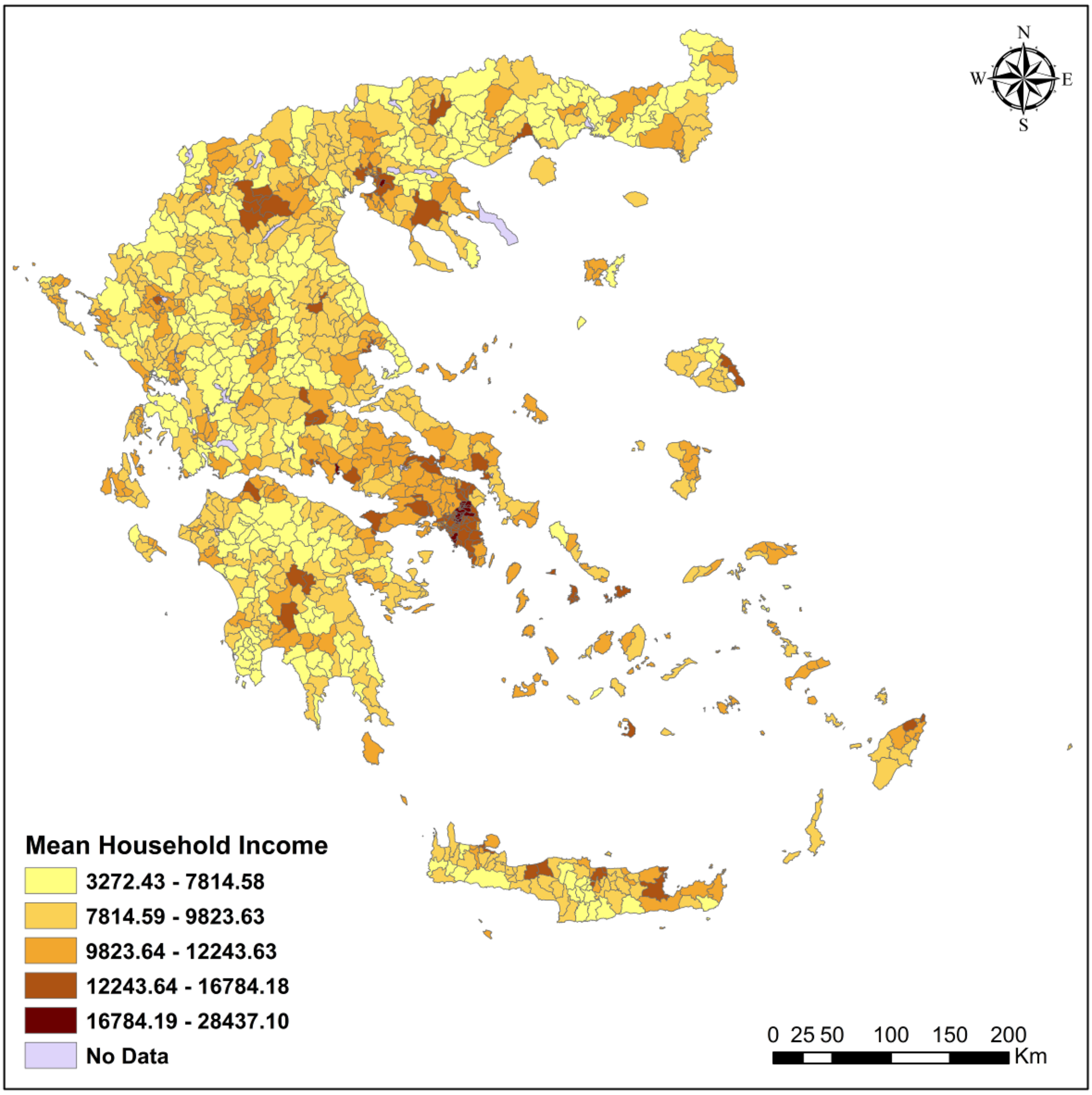

Figure 2. Mean recorded household income for Local Authorities in Greece

Local Pearson Correlation Coefficients

The Pearson Correlation Coefficient r is a standard statistic for checking for multicollinearity in the independent variables of a linear regression model calibrated using Ordinary Least Squares (OLS). The formula to calculate r in order to examine the correlation between two variables X and Y that have a normal distribution, mean values of $\bar{x}$ and $\bar{y}$, and standard deviations of s_x and s_y, respectively, is:

$$r = \frac{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)}}\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{(n-1)}}} \quad (1)$$

where n is the number of observations. The equation can also be written as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

The coefficient *r* is statistically significant at a given significance level *a* if the absolute value of *t* given by equation 3 is higher than the value of the two tailed t-student distribution for *n-2* degrees of freedom and significance level *a*. The formula of *t* is:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \Rightarrow t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (3)$$

where *n* is the number of observations (Kalogirou 2012).

The Local Pearson Correlation Coefficient (LPCC) for each observation point *i* is calculated as follows:

$$r_i = \frac{\sum_{j=1}^{k}(x_j - \bar{x}_i)(y_j - \bar{y}_i)}{\sqrt{\sum_{j=1}^{k}(x_j - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{k}(y_j - \bar{y}_i)^2}} \quad (4)$$

where *k* is the number of nearest neighbours around observation point *i*, $\bar{x}_i$ and $\bar{y}_i$ are the mean values of *x*s and *y*s of the *k* nearest neighbours of *i*. For example, $\bar{x}_i = \sum_{j=1}^{k} x_j \big/ k$.

Geographically Weighted Pearson Correlation

The GWPCC is a geographically weighted moment-based statistic that adopts the idea of geographical weighting of the values around an observation for which local statistics are calculated [8]. The formula to calculate $gwpcc_i$ in order to examine for local correlation between two variables *X* and *Y* that have a normal distribution, geographically weighted mean values of $\bar{x}_i = \sum_{j=1}^{k} x_j w_{ij} \big/ \sum_{j=1}^{k} w_{ij}$ and $\bar{y}_i = \sum_{j=1}^{k} y_j w_{ij} \big/ \sum_{j=1}^{k} w_{ij}$, and geographically weighted standard deviations of s_{xi} and s_{yi}, respectively, is:

$$gwpcc_i = \frac{\sum_{j=1}^{k} w_{ij}(x_j - \bar{x}_i)(y_j - \bar{y}_i)}{\sqrt{\sum_{j=1}^{k} w_{ij}(x_j - \bar{x}_i)^2}\sqrt{\sum_{j=1}^{k} w_{ij}(y_j - \bar{y}_i)^2}} \quad (5)$$

The function to calculate the weights w_{ij} in equation 5 is the bi-square function (Equation 6) also used in the literature for the adaptive kernel weighting scheme in GWR (Fotheringham et al. 2002).

$$w_{ij} = \begin{cases} [1-(d_{ij}/h_i)^2]^2 & \text{if } d_{ij} \le h_i \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where for each point i in which a local correlation coefficient is calculated, the weight in the point j is w_{ij}, d_{ij} being the distance between i and j, and h_i being the distance of the Nth nearest neighbour of i from the location of i ($h_i=d_{iN}$).

The number of nearest neighbours defines the spatial kernel and could thus be referred to as *bandwidth*. The latter can be set to a fixed value or allowed to vary. Here, LPCCs and GWPCCs have been calculated for 100 different bandwidths and are presented in Section 4.

4. Results and Discussion

The global Pearson correlation coefficient between the proportion of foreign citizens and income at local authority level in Greece is 0.2826 and it is statistically significant at the 99.99% level. Although a significant positive relationship between the two variables is found, the correlation coefficient magnitude suggests a rather weak relationship. However, this conclusion is misleading if this relationship is looked locally. To test for this, local correlation coefficients were calculated for 100 different bandwidths that are defined as 1% accumulative intervals (1-100%) of the total observations (1033 local authorities in Greece).

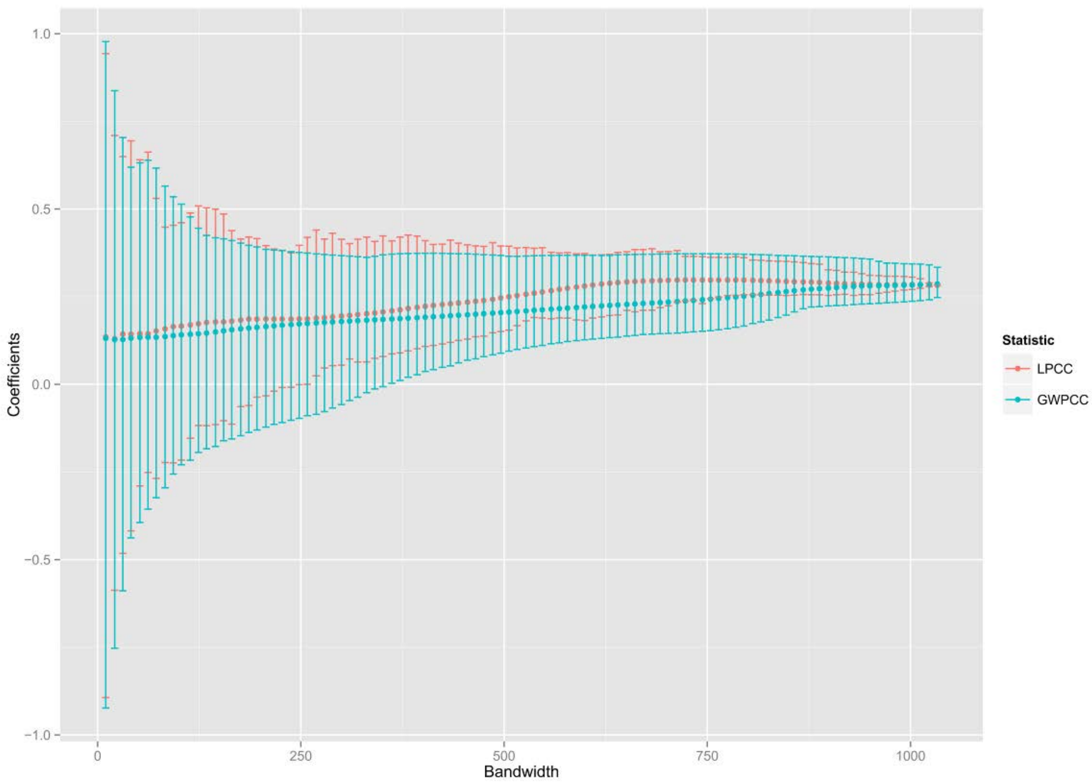

Figure 3. Local correlation coefficients for different bandwidths

Boxplots of the resulted LPCCs and GWPCCs along with mean values are presented in Figure 3. It is apparent from the latter that the smaller the kernel size, the wider the range of the local coefficients, which is an expected finding. Clearly, there is a spatial variation in the connection between the two variables. For example, for a bandwidth of 10 nearest neighbours LPCCs and GWPCCs range from -0.894 to 0.943 and -0.923 to 0.977, respectively.

In order to study the spatial variation of the local correlation coefficients and to assess if this variation is statistically significant original and simulated statistics were computed. LPCCs and GWPCCs were calculated for a fixed bandwidth (n=50) for the original locations of the observations as well as for 99 random geographical reallocations of the observations. The latter is a simple Monte Carlo simulation proposed by Hope (1968) and adopted by Fotheringham et al. (2002) who assess if local parameter estimates in a Geographically Weighted Regression model exhibit spatial non-stationarity.

Similarly to Fotheringham et al. (2002), the variances of LPCCs and GWPCCs, respectively, have been computed for observed and simulated local correlation coefficients. The variance of the observed LPCCs for a bandwidth n = 50 nearest neighbours is 0.0331 while the variances of the 99 simulated data sets range from 0.0081 to 0.0246. Since the variance of the observed data set is higher than all sets of simulated data, based on Hope's simple Monte Carlo test (1968), it can be concluded that the LPCCs exhibit significant spatial variation at the 99% level of significance. The variance of the observed GWPCCs for a bandwidth of 50 nearest neighbours is 0.0401 while the variances of the 99 simulated data sets range from 0.0174 to 0.0368. Thus, the GWPCCs also exhibit significant spatial variation at the 99% level of significance. Figure 6 shows a graph of the variances and the range of observed (Iteration 100) and simulated (Iterations 1-99) LPCCs and GWPCCs for the Monte Carol test.

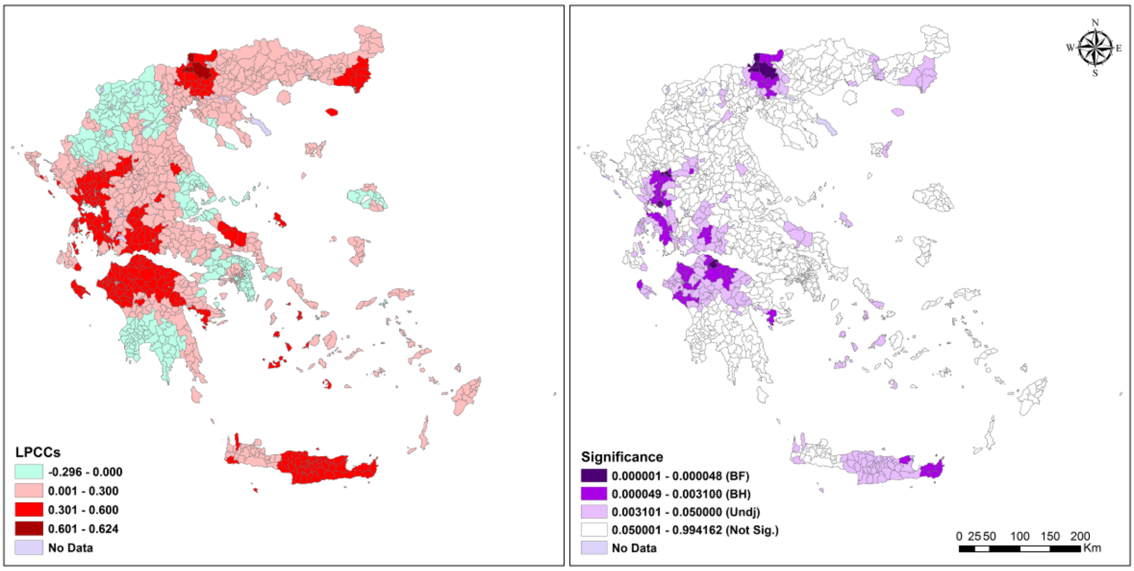

Figure 4. LPCCs for foreigners and mean household income

Figures 4 and 5 show maps of the LPCCs and GWPCCs. Maps of significance levels are also presented after correcting for multiple hypotheses testing. In Figures 4 and 5, BF stands for the Bonferroni (1935) correction and BH for the Benjamini and Hochberg (1995) correction.

The spatial patterns suggest significantly positive relationship between foreigners and income in local authorities with touristic and agricultural based local economies located in Crete, Cyclades, Ionian Islands, North Peloponnese, West Greece and Central Macedonia. Urban centres in Athens and Thessaloniki Metropolitan areas suggest a weak relationship occasionally positive or negative. A moderate significantly negative relationship appears in rural areas in West Macedonia and West Attica. The latter may be due to affordable housing near Metropolitan areas where these foreigners may be working.

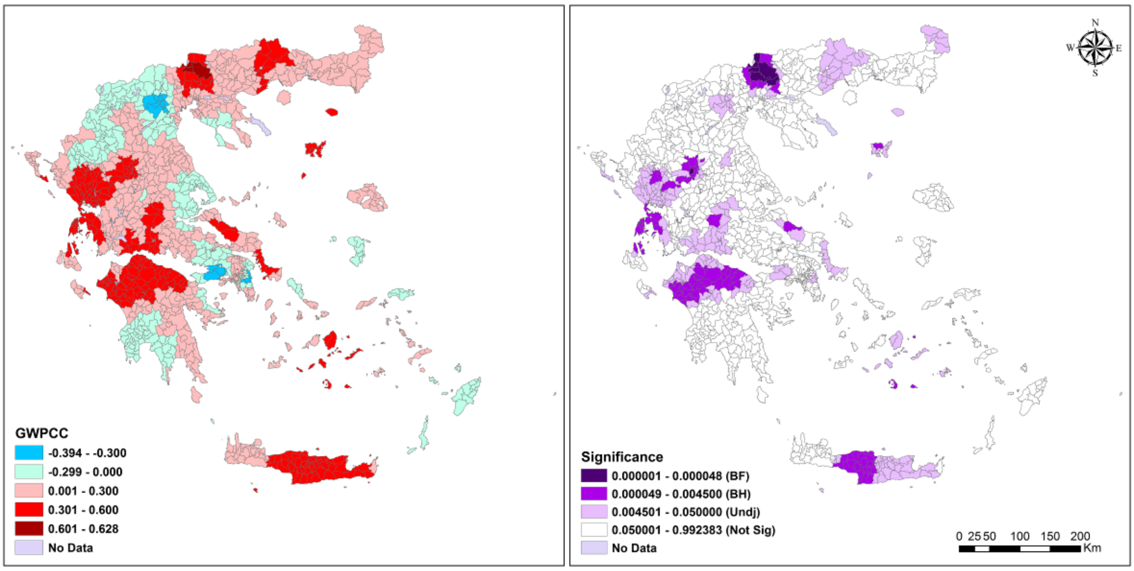

Figure 5. GWPCCs for foreigners and mean household income

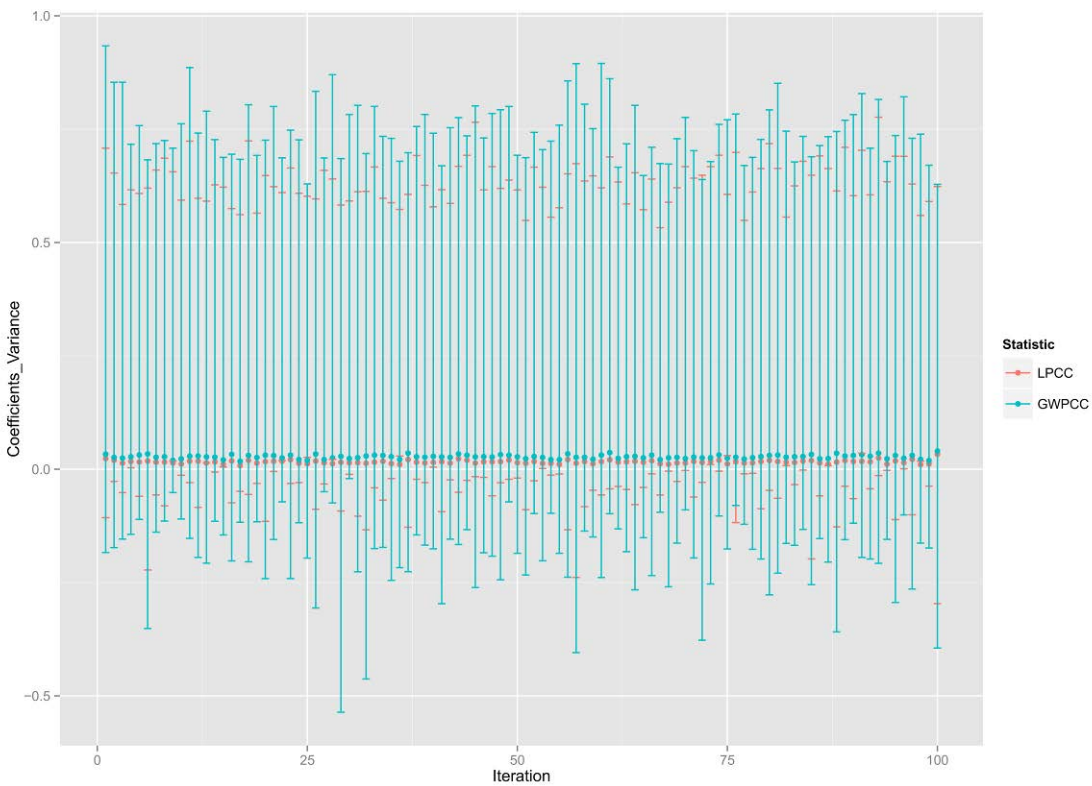

Figure 6. Observed and simulated LPCCs and GWPCCs

5. Conclusions

This paper provides empirical evidence that the relationship between the proportion of foreign citizens and household income at local authority level in Greece exhibits significant spatial variation. The global Pearson correlation coefficient suggested a positive and significant relationship. However, after calculating local correlation coefficients, it appeared that this relationship is significantly positive (and stronger) in local authorities with agricultural and touristic local economies. This may be due to the fact that since the majority of foreigners are economic immigrants they tend to reside in areas where higher incomes are generated. The latter is not the case in Athens and Thessaloniki Metropolitan areas where a weak and variable relationship is evident.

6. Acknowledgements

The development of the R package lctools and this research took place during my research visit in the Institute of Geography and Spatial Planning, University of Luxembourg in 2013/14. Therefore, I would like to sincerely thank all staff of the Institute of Geography and Spatial Planning and the University of Luxembourg for providing me with a nice working environment and to specially thank Prof. Geoffrey Caruso for his support.

7. References

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal Statistical Society Series B, 57, 289–300.

Bonferroni, C.E., 1935. Il calcolo delle assicurazioni su gruppi di teste, In *Studi in Onore del Professore Salvatore Ortu Carboni,* pp. 13-60, Rome, Italy.

Fotheringham, A.S, Brunsdon, C. and Charlton, M., 2002. Geographically Weighted Regression: the analysis of spatially varying relationships. John Wiley and Sons, Chichester.

Hope, A.C.A., 1968. A Simplified Monte Carlo Significance Test Procedure, Journal of the Royal Statistical Society. Series B (Methodological), 30 (3), 582 – 598.

Kalogirou, S., 2010. Spatial inequalities in income and post-graduate educational attainment in Greece. Journal of Maps, 6(1), 393–400. doi: 10.4113/jom.2010.1095

Kalogirou, S., 2012. Testing local versions of correlation coefficients, Review of Regional Research - Jahrbuch für Regionalwissenschaft, 32(1), 45 – 61. doi: 10.1007/s10037-011-0061-y

Kalogirou, S., 2015. lctools 0.2-2: Local Correlation, Spatial Inequalities and Other Tools, http://cran.r-project.org/web/packages/lctools/index.html

Kalogirou, S. and Hatzichristos, T., 2007. A spatial modelling framework for income estimation, Spatial Economic Analysis, 2(3), 297–316. doi: 10.1080/17421770701576921

Maloutas, T. and Karadimitriou N., 2001. Vertical social differentiation in Athens: alternative or complement to community segregation? International Journal of Urban and Regional Research, 25(4), 699–716. doi: 10.1111/1468-2427.00340

Rovolis, A. and Tragaki, A., 2006. Ethnic Characteristics and Geographical Distribution of Immigrants in Greece. European Urban and Regional Studies, 13 (2), 99-111. doi: 10.1177/0969776406062521

Rovolis, A. and Tragaki, A., 2008. The Regional Dimension of Migration in Greece, In: Coccossis, H., Psycharis, Y. (eds) Regional Analysis and Policy, pp. 99-117, Physica-Verlag, Heidelberg.