Spatial Pattern in Vacancy Rates
in English and Welsh Retail Centers

1. Introduction

Great Recession (GR) refers to worldwide-scale of economic recession resulted by Global Financial Crisis (GFC) from mid 2007 to early 2009, and it is one of the most dreadful economic crises in history. International Monetary Fund (IMF) asserts it is the worst economic recession after the Great Depression in 1930's (IMF, 2009). United Kingdom (UK) also severely suffered from GR. For UK, the period of GR is between 2nd quarter of 2008 and 2nd quarter of 2009. During this period of time, the average of quarterly growth rate in Gross Domestic Product (GDP) per capita, and household final consumption expenditure were -1.40% and -0.7%, respectively. Evidently, GR was a grim blow to UK's economy. Due to the recession, a lot of businesses were forced to close their branches and stores. According to an article from Telegraph (Telegraph, 2009), nearly 27,000 businesses closed during the recession.

A question rises whether there is a systematic pattern in store closures in terms of space. For instance, on a national level, have some cities in UK statistically significant higher or lower numbers of empty stores in their retail centers than other cities? In a regional level, are some areas in a region has more or less vacancy rate than other areas?

This report focuses on aforementioned question in two different levels, national level and regional level, and answers three distinct academic questions. First question is a general question: is there any spatial pattern in difference in vacancy rates before and after GFC in English and Welsh retail centers? Away from national level, the report seeks an answer for the same question as the former but on a smaller regional scale. Lastly, I test if certain factors, income, crime rate, and etc., play a big role in vacancy rates in Greater London area.

2. Data

Two significant data files drive the main results of this report. The first is a file named *Town Centre Statistics*. The file consists of 2 tables and each table contains 1907 observations of retail centers, and 8 variables describing each retail center in 2009 and 2014, respectively. The 8 variables are Department for Communities and Local Government[1] (DCLG) retail center ID, DCLG retail center name, total number of premises, proportion of comparison premises, proportion of convenience premises, proportion of leisure premises, proportion of service premises, proportion of vacant premises. The other file is a shape file which shows where each DCLG retail centers are located. The file shows the locations of 1312 retail centers as points and it is based on Ordnance Survey National Grid reference system, especially on Ordnance Survey Great Britain 1936 (OSGB36). Both of the files only focus on towns in England and Wales.

Main focus of this report is finding any existent clusters in terms of vacancy differences[2] between 2009 and 2014. After removing any retail centers with missing values in proportion of vacant premises column, 1259 retail centers remain in the data set.

For the analysis related to the third academic question, I use additional two data files from Consumer Data Research Centre[3] (CDRC): *CDRC 2015 English indices of deprivation Geodata Pack: London (E12000007)*, and *CDRC 1995-2016 House Prices Geodata Pack: London (E12000007)*. Both of the data files contain geo-information of Greater London area in multipolygon format and relative statistics in Greater London area at the Lower Layer Super Output Area[4] (LSOA) level.

The former shows which area is deprived in seven categories and a compound index made by those seven categories: the Index of Multiple Deprivation (IMD). The index is constructed mostly by statistics from 2012/13. The file contains both rank and decile information of all 8 variables, but I only use decile columns for my analysis. The latter contains information of the annual median transaction values per property during the period 1995-2016. To maintain consistency with data collection period in IMD data file, I only retrieve the data points from 2013. More detailed explanation is in section 3.3.

---

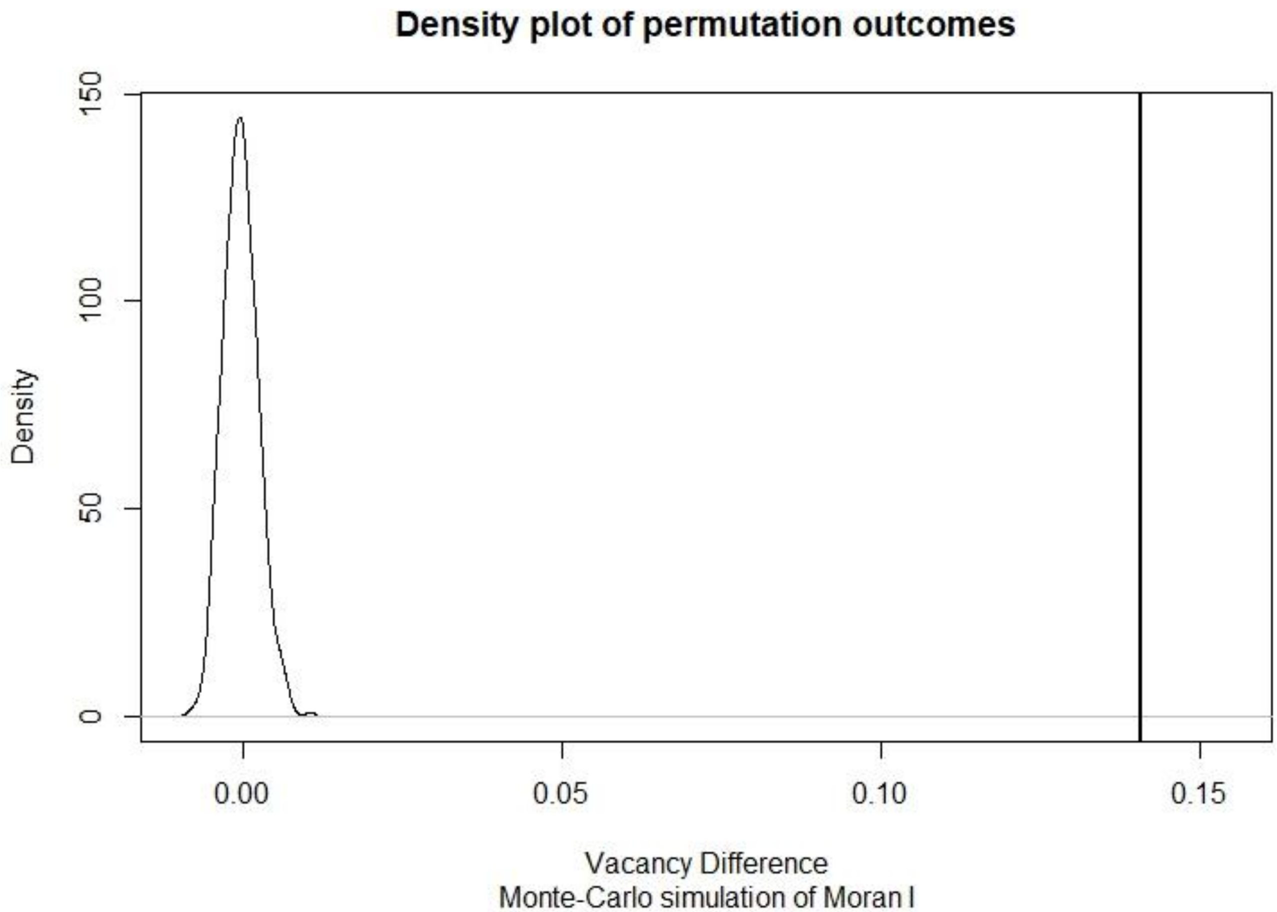
1 Currently, The Ministry of Housing, Communities and Local Government

2 I define vacancy difference(s) as the difference(s) between 2009 values and 2014 values of proportion of vacant premises in a(the) retail center(s).

3 cdrc.ac.uk/

4 For more specific information, please refer to https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

Figure 1

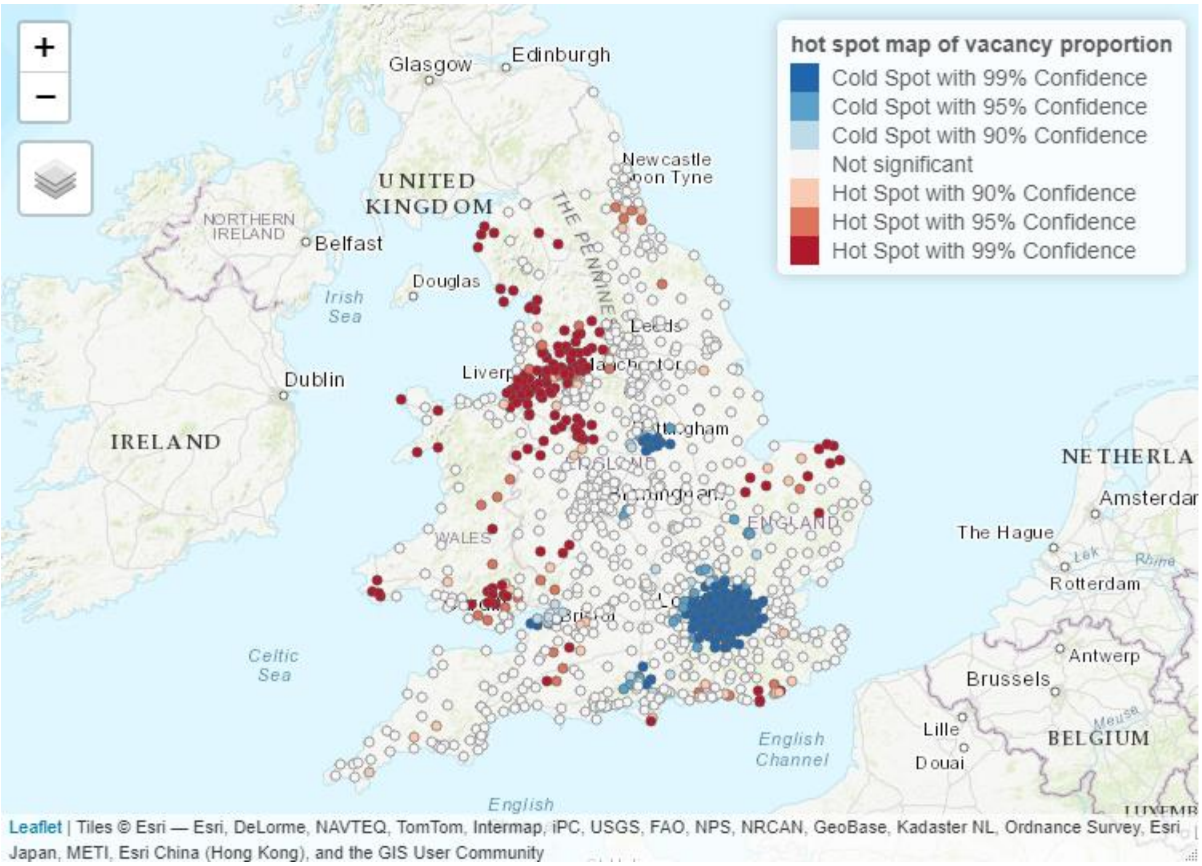


3. Analysis

3.1 England and Wales

In this subsection, I explain a spatial structure I have found in differences in proportion of vacant premises between 2009 and 2014 on a national scale. To verify whether a spatial structure exists on a national scale, I have calculated global Moran's *I* with the dataset. Moran's *I* has been firstly introduced by Patrick Alfred Pierce Moran in 1950 (Moran, 1950). Moran's *I* detects whether a spatial autocorrelation exists, meaning a given spatial structure is not random. It is similar to Pearson's correlation coefficient but applied in spatial sense. A researcher should be cautious in determining of statistical importance of Moran's *I* as it depends on the dataset whether the value of Moran's *I* is statistically significant or not. Also, Moran's *I* is dependent on spatial weights matrix hence spatial weights matrix must be carefully designed.

Figure 2 Hotspot Analysis on England and Wales Retail Centers



I have calculated spatial weights as follows. Neighborhood is defined by distance as the shape file of DCLG retail centers is comprised of spatial points. More specifically, if the distance between two different points is less than 50km, they are neighbors to each other. Then, I applied Double-power distance weighting scheme to generate the spatial weights matrix.

The global Moran's *I* has been derived from the difference between 2009 and 2014 in proportion of vacant premises in each retail centers and the spatial weights matrix I've constructed. Figure 1 shows the distribution of Moran's *I* values if it is simulated in Monte-Carlo fashion by 1000 times and the straight line on the right side of the graph depicts the calculated global Moran's *I* value from the dataset. The value of global Moran's *I* is 0.141 if I round it up at third decimal point and it is statistically highly significant. Low value of its p-value (0.000999) supports the fact the given spatial structure in the dataset is not random but spatially correlated.

Note that, the focus of this report is not to provide explanation what drives the spatial structure but rather to show there is an underlying spatial structure related to vacancy differences in each retail centers and, furthermore, present spatial clusters if any of them exists.

I conduct a Hotspot analysis by using ArcGIS. Hotspot analysis is based on Getis-Ord statistic, or commonly called G-statistic[5]. Figure 2 illustrates the result of the Hotspot analysis. White points are points with no statistically significant values in G-statistic. Points with red points have positive G-statistic with 99% confidence; orange points 95%; light orange points 90%. On the other hand, blue dots have negative values in G-statistic with 1% significance level; light blue dots 5%; very light blue dots 10%.

Distinctive patter exists. Areas surrounding Liverpool and Manchester have higher values in vacancy differences. On the contrary, areas around London have lower values, meaning there were less vacant premises in retail centers near to London in 2014 than 2009. Moreover, areas around Cardiff and several retail centers near the coastline share similar pattern as areas near to Liverpool and Manchester. Retail centers near to Nottingham and a few retail centers near Bristol and Southampton have lower values in difference in vacancy rates.

3.2 Top 3 Metropolitan Areas

After searching for spatial pattern on national scale, I have narrowed my scope to three biggest metropolitan area in England and Wales in terms of population size[6]. The three areas are Greater London (London), West Midlands (Birmingham), and Greater Manchester (Manchester). My initial guess was retail centers close to a central business district (CBD) of a Metropolitan area would have lower values in vacancy differences (cold spots), and suburban areas would have higher values (hot spots). The reasoning of my assumption has a lot in common with Alonso's land use model (1964). As CBD is the pinnacle of commercial area in metropolitan area, I presume it provides more customers, even the wealthy ones, which would lead to more profits for businesses near to CBD. Thus, the retail centers which are close to CBD would, at least, have lower values in vacancy differences than the retail centers in suburban area.

---

5 Full table of g-statistics used in my Hotspot analysis could be provided upon request.
6 It is based on 2011 census data.

Figure 3 Hotspot Analysis on Greater London

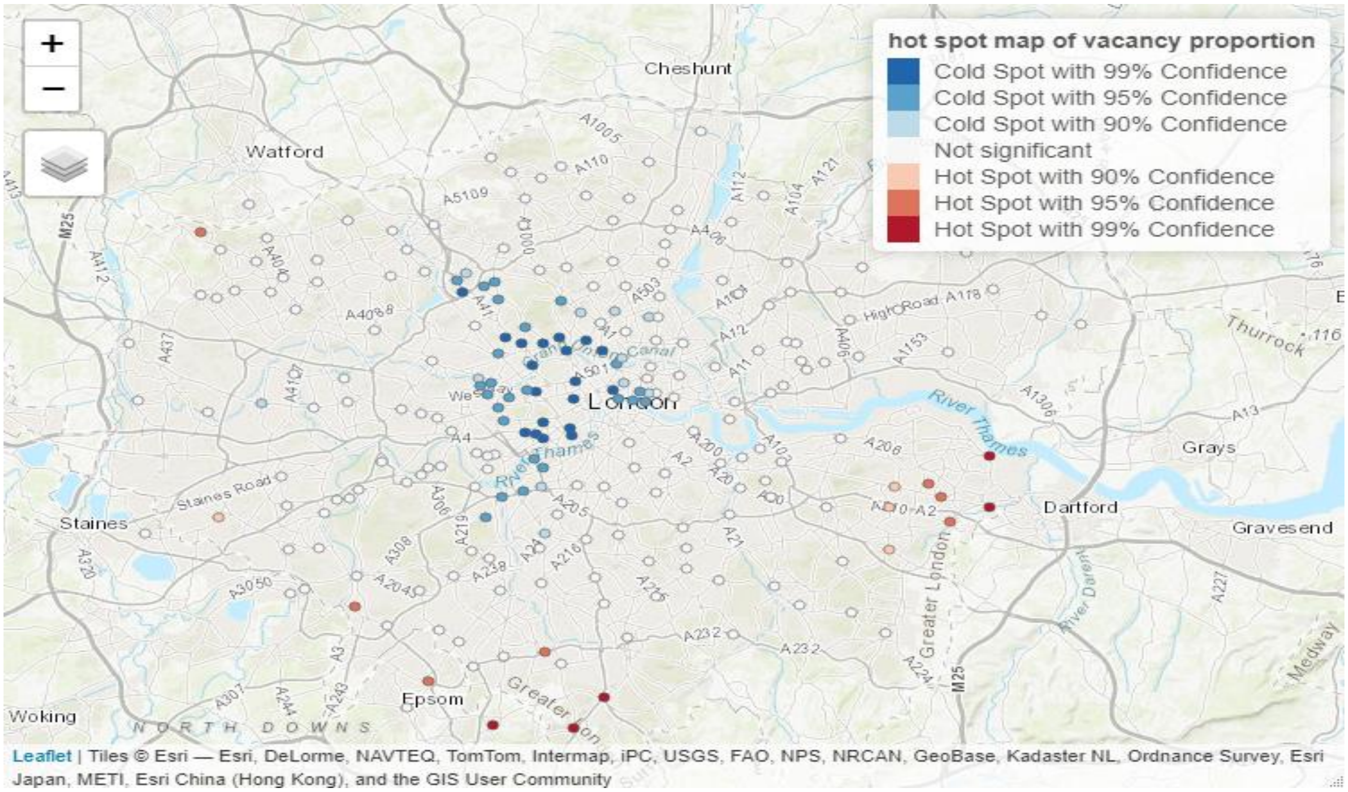


Figure 4 Hotspot Analysis on Greater Manchester

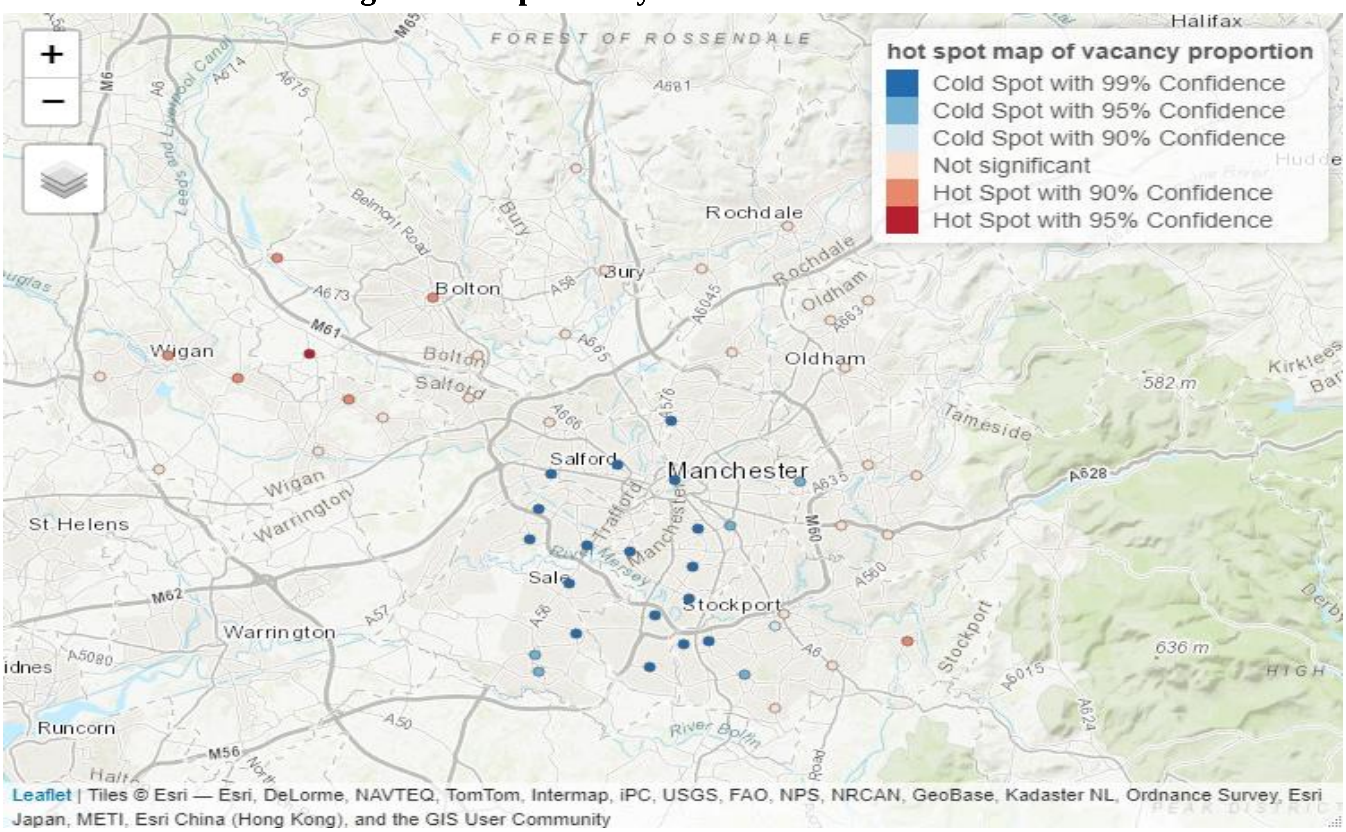
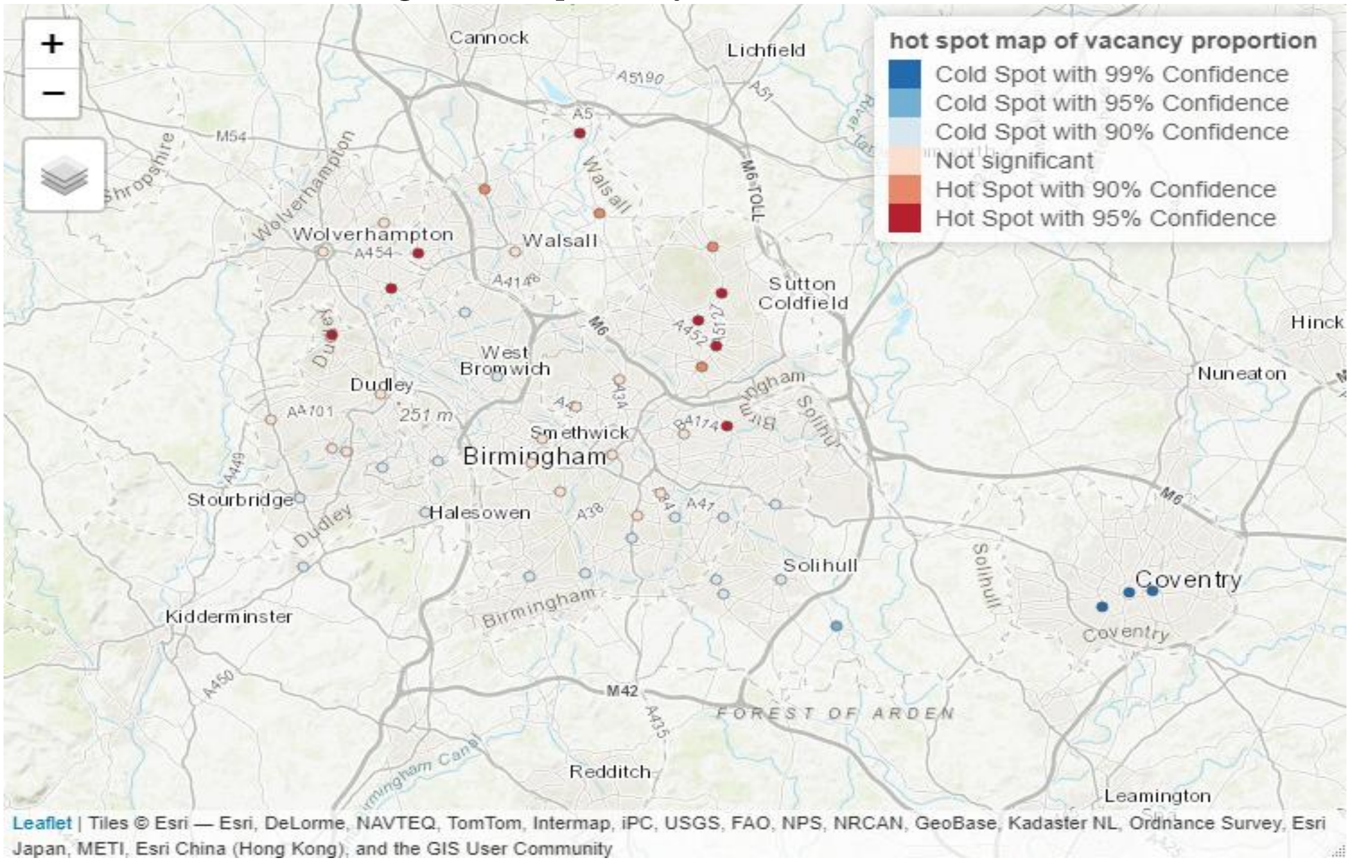


Figure 3 Hotspot Analysis on West Midlands



The result of Hotspot analysis in three metropolitan areas are shown in Figure 3,4 and 5. Greater London and Greater Manchester areas show a spatial pattern as I have expected. Retail centers near to each main cities' CBD form a cluster of cold spots: less vacant premises after the GR, and some retail centers located on the periphery of a metropolitan area have statistically higher vacancy differences than other retail centers.

Interestingly, West Midlands, however, show a different pattern. 3 retail centers in Coventry have significantly lower vacancy differences than other retail centers around Birmingham. Retail centers in the northern parts of West Midlands, do have higher vacancy differences than other retail centers but the retail centers in CBD of Birmingham do not have vacancy differences which are on the either end of the distribution. The structure is partly from the fact the retail centers in Coventry are Cold spots with 90% significance even on the national scale.

3.3 Greater London

In this subsection, I explore if there is any factor which would have significant explanatory value in a retail center's vacancy difference between 2009 and 2014. Firstly, I used the data pack, *CDRC 2015*

Table 1 OLS Results

	(1) Coefficients	(2) Coefficients	
Intercept	0.0097	0.0155	
	(0.0164)	(0.0112)	
Income	-0.0067	-0.0057	**
	(0.005)	(0.0028)	
Health	0.0032	0.0035	
	(0.0031)	(0.0029)	
Living Environment	0.0049	0.0046	*
	(0.0031)	(0.0027)	
Housing	-0.0004		
	(0.0026)		
Employment	0.0005		
	(0.0047)		
Education	0.0013		
	(0.0028)		
Crime	0.0003		
	(0.0026)		
Adjusted R^2	0.0082	0.0251	
F-statistic	1.264	2.911	**

Standard errors are in parentheses. * stands for 10% significance level; ** is for 5% significance level; *** means 1% significance level.

English indices of deprivation Geodata Pack: London, for the analysis. Index of Multiple Deprivation (IMD) constitutes of 7 sub-indices: Income Deprivation; Employment Deprivation; Education, Skills and Training Deprivation; Health Deprivation and Disability; Crime; Barriers to Housing and Services; and Living Environment Deprivation. Income Deprivation shows the proportion of a LSOA population who have “low-income”. “low-income” refers to both people with no jobs and with jobs but low salary. Employment Deprivation calculates the share of a LSOA population who are in working age, from 16 to 64, are willing to work but currently not active in labor market. Education, Skills and Training Deprivation measures how population of a LSOA is deprived of educational attainment and skills. Health Deprivation and Disability illustrates morbidity, disability, and premature mortality of a LSOA and does not take behavioral and environmental risks into account. Crime sub-index calculates the risk of a person in LSOA being a victim of a crime. Barriers to Housing and Services measures how population in a LSOA is deprived of housing and local services due to financial and geographical barriers. Living Environment Deprivation evaluates the quality of both "indoors" and "outdoors" living environment of a LSOA. "indoors" living environment is of the quality of housing and "outdoors" is of air quality and risks of getting into a traffic accident. For more detailed description of how the IMD is calculated, please refer to Department for Communities and Local Government (2015).

To verify whether these seven sub-indices have statistically significant values in explaining a retail center's vacancy difference, I conduct a simple OLS with following equation.

$$\boldsymbol{Vacancy\ Difference_i = \beta_1 Income_i + \beta_2 Health_i + \beta_3 Living\ Environment_i + \beta_4 Housing_i + \beta_5 Employment_i + \beta_6 Education_i + \beta_7 Crime_i + \varepsilon_i} \qquad \mathbf{(1)}$$

The result of OLS analysis of equation (1) is shown in first column of Table 1. None of the sub-indices are statistically significant and the analysis results in low values in adjusted R^2 and F-statistic. Due to the low values, the model itself can provide little explanation in variation of vacancy differences in retail centers and the null hypothesis, all the coefficients of the parameters are zero, cannot be rejected.

As I was convinced at least income should carry some explanatory value, I concluded some parameters are redundant and are harming the model. Thus, I reran OLS with fewer independent variables which had higher t-values than 1. The new regression equation is as below.

$$\boldsymbol{Vacancy\ Difference_i = \beta_1 Income_i + \beta_2 Health_i + \beta_3 Living\ Environment_i + \varepsilon_i} \qquad \mathbf{(2)}$$

The regression result of equation (2) is displayed in second column of Table 1. As I expected, adjusted R^2 and F-statistic increased significantly. The new model does have explanatory value and the null hypothesis that all betas are zero could be rejected. Moreover, both *Income* variable and *Living Environment* variable are statistically significant, at 5% level and 10% level respectively.

If the model is true, a vacancy difference of a LSOA would decrease by 0.0057(0.57% points) if the income decile of the LSOA has increased by 1[7]. For instance, if a LSOA belongs the lowest 10%

7 Note that each independent variable is a categorical variable which ranges from 1 to 10, all integers, and 1 means a LSOA belongs to lowest 10% decile in corresponding category.

income decile among other LSOA in Greater London but has become to be in between the lowest 10% and the lowest 20% income decile, the vacancy difference would decrease by 0.57 percentage points. Applying the same logic, vacancy difference of a LSOA would increase by 0.46 percentage points if the *Living Environment* variable has increased by 1, which is counterintuitive. As it is only statistically significant at 10% level, I disregard its importance.

To seek whether space affects in vacancy differences, I calculated the Moran's I with the residuals of both OLS regressions. Each Moran's *I* is permutated 1000 times with Monte-Carlo style and both values are insignificant: both p-values are 0.4256 and 0.4855. However, I believed some spatial autocorrelation must exist related to property price. Thus, I construct an extended model of equation (2) just adding additional variable *House Price* which I have acquired from *CDRC 1995-2016 House Prices Geodata Pack: London (E12000007)* and is of annual median property price in a LSOA. Then, I calculated Moran's *I* with its residuals. Nevertheless, the obtained value is still statistically insignificant. The p-value is 0.4525. Hence, I concluded, at least in Greater London area, no spatial autocorrelation exists in terms of vacancy differences, and income decile is highly negatively correlated with vacancy differences.

4. Discussion and Conclusion

Readers should note several shortcomings of this report. Firstly, this report simply shows spatial clusters derived from a GIS program and does not provide plausible explanation behind it. A reasonable explanation is necessary to support the existence of spatial clusters. Secondly, the data acquisition period is not appropriate of the analysis. The Great Recession (GR) is from 2nd quarter of 2008 to 2nd quarter of 2009. However, the main dependent variable is the difference of vacancy rates between 2009 and 2014. Thus, it is plausible to think main dependent variable does not have values before the GR to analyze its true effect on vacancy differences, different from the description of the data pack. It would be more suitable if a researcher could compare values in 2014 with values in year before 2008 and analyze the impact of GR on vacancy differences of retail centers.

This report aims to reveal any spatial structures in terms of vacancy differences in English and Welsh retail centers, especially clusters. It has shown spatial clusters exist both on national and regional scale. However, the report does not provide any explanation why such these clusters form and it is out of the scope of this report. However, at least in Greater London, the report explains vacancy difference of a retail center is highly negatively correlated with income deciles of surrounding LSOA and the hypothesis that spatial autocorrelation does not exist cannot be rejected.

Word count: 2681

References

1. Alonso, W. 1964. *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, MA: Harvard University Press.
2. Department for Communities and Local Government. September 30, 2015. *The English Indices of Deprivation 2015* (PDF)
3. IMF. April 24, 2009. *World Economic Outlook — April 2009: Crisis and Recovery* (PDF). *Box 1.1 (page 11-14)*. Retrieved September 17, 2013.
4. Moran, P.A. 1950. *Notes on continuous stochastic phenomena*. Biometrika, 37(1/2), pp.17-23.
5. Telegraph. December 23, 2009. *UK recession the worst for company failures as nearly 27,000 go bust, says Conservatives*