

Zofia Bednarowska

Jagiellonian University in Kraków
e-mail: zofia.bednarowska@uj.edu.pl

**A SPATIAL REGRESSION MODEL
OF RETAIL CHAINS DEVELOPMENT IN POLAND**

**MODEL REGRESJI PRZESTRZENNEJ ROZWOJU
SIECI HANDLOWYCH W POLSCE**

DOI: 10.15611/ekt.2015.3.04

Summary: The key research goal is to identify predictors of retail chain space development in Poland, as well as to define if there is any spatial correlation between increasing retail area and the spatial proximity of the other malls. In order to do so, the study covers an analysis of the retail area in Poland in the framework of the socio-economic development of cities. The study covers a macroeconomic overview of the retail map of Poland. The unit of the analysis are cities with a population of 40,000-400,000 inhabitants. The method of analysis is based on two models: non-spatial OLS and the spatial regression model. The model's goal is to predict the retail development of a city based on the most effective predictors. The independent variables included a range of socio-economic indicators such as: city population, city unemployment rate and average salary in the private sector.

Keywords: spatial econometrics (statistics), urban space, retail chain development.

Streszczenie: Celem badania było zidentyfikowanie czynników odpowiedzialnych za rozwój sieci handlu detalicznego, a także zdiagnozowanie, czy istnieje przestrzenna korelacja między rozwojem tych sieci a ich wzajemną przestrzenną bliskością w różnych miastach. W tym celu przeprowadzono analizę statystyczną przestrzeni handlowej w Polsce uwzględniającą rozwój społeczno-ekonomiczny w miastach o populacji 40 000-400 000 mieszkańców. Metoda analizy opierała się na modelu regresji liniowej i regresji przestrzennej.

Słowa kluczowe: ekonometria przestrzenna, przestrzeń konsumpcji, rozwój sieci handlu detalicznego

*Everything is related
to everything else, but near things
are more related than distant things.*

Tobler's First Law of Geography

1. Introduction

The main focus of the research problem explores the phenomenon of increasing consumption space within urban areas. The significance of consumption is growing, particularly in urban areas, to the extent that we witness living in consumer cities [Miles 2012, p. 2].

One can observe a decreasing level of the “exploitation” of existing commercial centres (shopping malls), while paradoxically, new larger commercial buildings are being constructed in nearby areas. What is more, some so-called first generation commercial centres are said to be dying. Dead malls have become an increasingly familiar image of urban and suburban space, largely due to the oversaturation of the market [Schwartz 2015]. This is only beginning in Europe, yet dead malls are well known in the USA and China. The question is to what extent does consumption society need more consumption space, and where will these trends lead us?

At the same time, anchor stores (usually they are part of retail chains) are developing. They are either neighbours of shopping malls or standing on their own. However, their development and survival right next to dying malls is even more interesting.

The development of new commercial space (here defined as shopping malls and anchor stores) and the death of old ones is reflected in the spatial and social transformation of cities. The function of the city centre narrows to a purely shopping and service area, shopping malls often being situated all over the city become the focal points of the districts, which results in the uneven development of city districts. The research project is aimed at exploring factors influencing spatial development. Firstly, from a macro-sociological perspective – transformation is the result of the economic condition of the area – its commercial potential (size of resident population, average salary, unemployment rate) and the demographic profile of its citizens. In addition to socio-economic factors, which explain part of the research problem, there are factors at the micro level which will be investigated (such as its significance to local communities' development), yet they are beyond the scope of this article.

The research questions that underlie this analysis are as follows: (1) To what extent is the development of consumption space commensurate with the potential of a city in Poland? (2) Is there any pattern behind the consumption area? Are nearby cities and towns similar to each other, e.g. small, oversaturated and located close to similar ones? (3) Does spatial correlation exist between retail chain areas, and if so, how does it contribute to a better explanation of the variance of increasing commercial space?

A possible answer to these questions will be provided by a non-spatial OLS regression model, as well as by a spatial regression model, aimed at explaining if there is a spatial correlation between the economic development of Polish regions and commercial space. It follows a similar approach applied in a statistical spatial analysis of economic and social development in Poland. It was proven that the spatial diversity of four stimulants (the gross value of fixed assets per capita, GDP per capita, investment per capita and salaries) and a destimulant of economic growth (unemployment rate) do not overlap [*Statystyczna analiza...* 2013, p. 31].

Many theories describe the above problem from various fragmentary perspectives, e.g. mention international trade, decision theories, retailer's international marketing policy, internationalization, market saturation, spatial interaction and central-place theory, oligopoly, and explaining the spatial expansion of particular chains. However, none of those explain the problem in terms of the socio-economic development of the city and its spatial proximity to other retail areas.

2. Spatial econometrics application

2.1. Spatial analysis

There are plenty of ways to introduce spatial analysis, which is of interest among many scientists. The high significance of maps as a spatial visual display of data is commonly acknowledged, although often it is limited to geostatistics and the exploratory spatial data analysis of the basic distribution implied on the basemap as a first layer. Therefore I would opt for a simple distinction between spatial analysis and spatial econometrics (statistics), while the former notion would be much too general, the latter would demand applying statistic tools. Suchecki [2010, p. 100] calls all of them as Exploratory Data Analysis (ESDA) which include also more developed process spatial econometrics.

Geographers categorize them a little differently: (1) Spatial Point Pattern Analysis, (2) Interpolation and Geostatistics, (3) Areal Data and Spatial Autocorrelation, including: Spatial Neighbours, Spatial Weights, Spatial Autocorrelation Tests and (4) Modelling Areal Data, including Spatial Statistics Approaches, Mixed-Effects Models, Spatial Econometrics Approaches and other methods [Bivand et al. 2008].

The importance of space was shown in a study of cholera by John Snow. Snow's maps of London have become models of how spatial correlation can embody causal thinking [Ward, Gleditsch 2008, p. 16]. I risk the statement that spatial econometrics has rarely been put under scrutiny in social and market research, unlike in Economics or Anglo-Saxon and American Political Science.

Only recently has a growing interest in spatial and spatial-temporal econometrics been seen in Poland – there were conferences organized by the University of Łódź, and afterwards examples of the analysis were published [Jewczak, Żółtaszek 2011].

There were interesting models applied, and some of them used spatial econometrics – spatial models of the professional activity rate and the unemployment rate in 25 EU countries, and the spatial differentiation of the mortality rate in Poland. There are plenty of contemporary topics covered by spatial statistics, such as the spatial relation between democracy and GDP or turnouts [Ward, Gleditsch 2008], second World War alliances [Franzese et al. 2012], economic dyads in trade, the gravity model of international trade, and house prices in the secondary market [Pietrzykowski 2011]. Below an example of such an analysis will be introduced.

I believe space to be a contributory variable in social and market research, due to the growing interest in space, which is visible in: geomapping, geocoding, geomarketing, geospatially oriented social media, geotagging, the development of Volunteered Geographic Information (VGI), and all of them are called "the geospatial revolution" [Robinson et al. 2015]. Since space is a key context of both consumer behaviour in B2C research and of companies' activity, distribution channel analysis etc. in B2B research, it shall be incorporated into the analyses.

The software available for spatial analysis for social scientists is pretty various: R (R-package *spdep*) and STATA (*spatreg* macro), MATLAB toolboxes, Splus, WINBUGS, set of macros for SAS [Ward, Gleditsch 2008 p. 87]. In the 23rd version of IBM SPSS there was the first introduction of "geospatial analytics". Other software options include GIS software (ESRI), e.g. ArcGIS, ArcGIS – Geostatistical Analyst, ArcGIS – Spatial Analyst, GeoDa™ (Anselin) and others, such as CrimeSTAT, MLwiN, TerraSeer (previous SpaceStat of Anselin), GeoVista Studio Project, IDRISI, QGIS, S+SpatialStats, Spatial Analysis for Macroecology, Spatial Analysis Utilities, STARS: Space Time Analysis of Regional Systems, Tobler's Flow Mapper [CSISS 2015].

2.2. Spatial econometrics (statistics)

Following Ward and Gleditsch [2008], I will introduce the basic assumptions of spatial econometrics. The key measures of spatial association and correlation are the spatial proximity of two observations i and j – S_{ij} and their similarity U_{ij} . A measure of spatial correlation may be various, but they recommend the Moran I statistic. The key notion is spatial lag – the scalar that sums the average across all neighbouring observations of one unit i , e.g. the average value of retail chains of neighbours. Ward and Gleditsch mention two types of models:

1. The spatially lagged y model (SLDV), where the values of y in one unit i are directly influenced by the values of y found in i 's "neighbours":

$$y_i = \beta_0 + \beta_1 x_i + \rho w_i \cdot y_i + \varepsilon_i$$

A positive value for the parameter associated with the spatial lag (ρ) means that observations have higher values if, on average, their neighbours have higher values. Variable w_i is crucial – "W refers to the matrix that weights the value of the spatially

lagged variable of other units. As unimportant as it may appear, W specifies, [...], why and how other units of analysis affect the unit under observation” [Neumayer, Plümper 2013].

2. In order to capture information about the structure of the error process ($\lambda w_i \cdot \zeta_i$) a second type of spatial-error model (SEM) is recommended:

$$y_i = \beta_0 + \beta_1 x_i + \lambda w_i \cdot \zeta_i + \varepsilon_i,$$

where: ε – a spatially uncorrelated error term; ζ – term of spatial component of error term; λ measures correlation of spatial component of the errors ε with one another for nearby observations [Ward, Gleditsch 2008, pp. 12-39].

In order to run a spatial regression analysis, a few procedures need to be performed such as mapping the data, i.e. specifying interdependencies among data – linkages assigned to the observations, which could be geographical (distances between centroids, length of borders between states) or economic closeness. There are more and more available GIS sources and country-level data. The key issue is properly choosing units of observations in the data set and transforming them to matrix (e.g. in R), since this is the necessary input. Afterwards, choosing and learning the software is crucial.

2.3. Model description

The data source is a market research data base of 105 Polish cities, provided by the market research agency report [PMR 2013], enriched by the author with geographical coordinates, recoded and transformed into a statistical data base. Observation with missing data was excluded listwise.

Prior to the analysis, a geostatistic basemap was introduced to see if there were any spatial relations between the density of the city population and number of commercial centres within a city. Green bubbles equals the number of commercial centres, heat choropleth equals density of population per 1 km². As one can see below, there are several similar neighbourhoods (clusters) – smaller cities and a smaller number of retail chains, as well as some cities with a big amount of chains, but a small population.

The unit of observation was a city from the category of 40,000-400,000 inhabitants, due to access to such data, so that the model is treated as a pilot model. After supplementing and updating a dataset, a full analysis will be completed. It is hard to define what kind of cities they are, due to many definitions of medium-sized cities. This follows the lack of sharp boundaries between small and medium-sized cities and results in methodological problems of researching medium-sized cities in Poland [Runge 2012, p. 85]. It is also said that townships are still the nationwide, regional and local growth centres [*Statystyczna analiza...* 2013, p. 63].

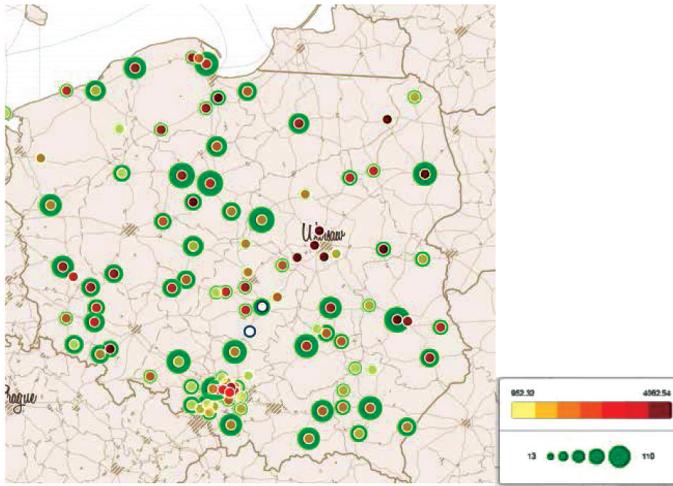


Figure 1. A map prepared within CartoDB Map software

Source: own calculations.

The method of analysis (modeling) is spatial econometrics, followed by the recommendation to run prior to a good non-spatial OLS regression model to modeling as a comparison [Franzese, Hays 2007].

The first non-spatial OLS regression model aimed to elaborate the best regression basic model and then compare the results. It was followed by the hypothesis that there is a relationship between the socio-economic development and spatial development of a retail area. The model's goal is to predict the retail development of the city based on the most effective predictors. There were several different predictors included. The independent variables covered a range of socio-economic indicators, such as: city population, city unemployment rate, average salary in the private sector. Other variables that were included in the first working model aimed at better conceptualizing of the economic development of the city (e.g. average disposable income, biggest firms rate), potential (city area in km², dwellings, planned shopping areas, planned number of shopping malls, number of companies operating in the city, as well as planned retail chains' areas). Some of them were overlapping in already provided information and explained variance, some remained non-significant, therefore they were excluded from the analysis. The dependent variable was the sum of retail chains within the city, both food and non-food¹.

Based on several attempts of a series of models, after the outliers exclusion, the most effective predictors were chosen since they were not easy to investigate and the results were not obvious. Many of the models had high R statistics, yet some of the

¹ It was planned to predict the number of shopping malls, but due to the small variation of the variable, it is more challenging to estimate. Further attempts are being continued.

predictors were non-significant. The final variables were standardized, and salary logged. Out of the previous models, model number 8 was chosen.

Table 1. Non-spatial OLS model statistics

| Model | R | R square | Adjusted R square | Mean Square | F | Sig. |
|-------|--------------|--------------|-------------------|-------------|---------|-------|
| 6 | 0.880 | 0.774 | 0.765 | 7,908.429 | 85,770 | 0.000 |
| 7 | 0.879 | 0.772 | 0.765 | 10,516.142 | 114,137 | 0.000 |
| 8 | 0.896 | 0.802 | 0.796 | 10,298.338 | 135,246 | 0.000 |

Source: own calculations.

When shifting to the spatial regression model, a list of the economic variables is enriched with the spatial autocorrelation in order to see if there is a relation between increasing retail chains and the spatial proximity of the other cities with chains. The model is as follows:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i + \beta_3 x_i + \rho w_i \cdot y_i + \varepsilon_i$$

where: y_i – sum of retail chains, β_0 – intercept, $\beta_1 x_i$ – population (standardized), $\beta_2 x_i$ – unemployment rate (standardized), $\beta_3 x_i$ – salary (logged), $\rho w_i \cdot y_i$ – spatial lag (ρ) and connectivity vector w_i , ε_i – error term.

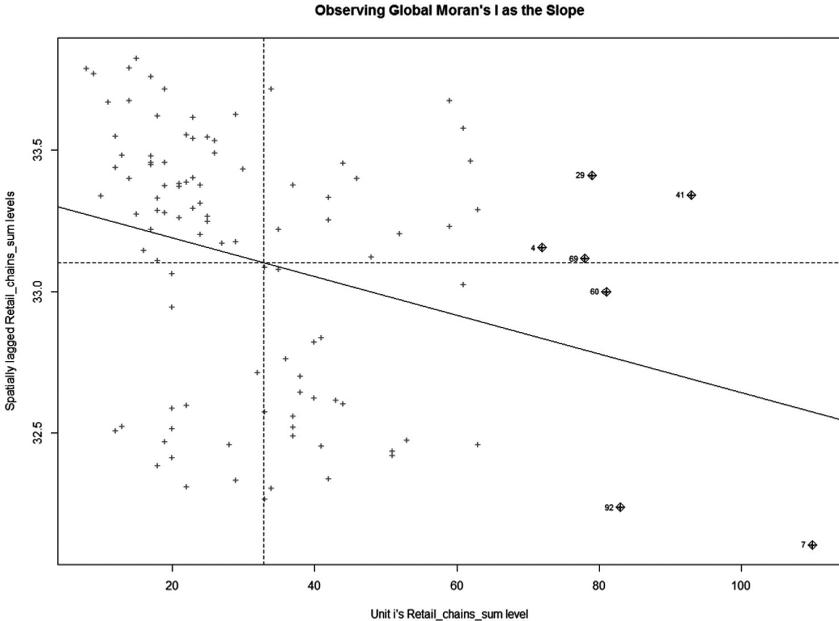


Figure 2. Moran’s scatterplot of a dependent variable and spatial lag

Source: plot generated by the author in R software.

The scatterplot of a dependent variable and its spatial lag below shows the model chosen. The plot shows in the lower right corner the points where there are a lot of chains inside the city but little in the neighbourhood; in the right above a lot of chains and a lot in the neighbourhood; in the upper left a lot in neighbourhood but a small number of chains; in the lower left a small number of both chains and neighbourhoods.

Following model statistics, it is non-significant – global Moran's I shows no spatial autocorrelation and as shown on the plot, the relations are not strong. If the model was significant, we could run spatial-error model (SEM). This would be possible because the Lagrange multiplier test of the second type – that is compared to the spatial-error alternative – is significant, therefore a connection between the w matrix and errors and split it up spatially by doing SEM.

2.4. Challenges and possible solutions

The results present the pilot model which is planned to be completed with data regarding cities with a population of over 400,000 people and updated values of key variables. In order to make the model more precise, a few improvements will be checked and introduced. If available in the new dataset, new variables that might be better predictors will be included into the non-spatial OLS model, or at least their other specification. In order to check model stability, a model on data from a different time period should be run.

Also helpful may be the rearrangement of the matrix, such as split data into two or more geographical areas. Matrices matter particularly in the case described in this paper, i.e. if the Lagrange Multiplier Test for residual autocorrelation (Rao's score test) opts for SEM, but intuition works rather for SLDV [Ward, Gleditsch 2008]. Due to the different structure of weight matrix, spatial correlations measures may change [Suchecki 2010, p. 112].

The matter of constructing and treating connectivities, even geographical ones, is always a challenge [Ward, Gleditsch 2008, p. 77]. Another option to consider is including geographical measures other than distances based on the longitude and latitude between particular cities (points). One option may be marking the consumption area of each particular city and its suburbs, which also contain some shopping malls, retail chain stores. That would mean shifting from an analysis based on points into geostatistical surface data [Ward, Gleditsch 2008, p. 84]. However, this results in detailed analytical and cartographic work prior to modeling.

Another option to be considered is including LISA – Local Indicators of Spatial Association [Anselin 1995]. An additional, yet much more challenging option would be incorporating nongeographical measures into the model, e.g. cultural similarity or economic distance [Beck et al. 2006].

3. Conclusions

The spatial OLS regression model contributes to a better understanding of the spatial retail transformation, interactions between cities, and directions of economic development. It is market potential that impacts retail chains' development – the population, average salary in the private sector and unemployment rate. The results may be applicable as a tool in assessing market business potential of particular cities and their neighbourhoods but also in public policy, smart city governance in order to increase the quality of life, and better spatial planning.

Furthermore, the SAR model aims at proving that the implementation of space in data analysis has great potential and that spatial analysis is contributory to a better understanding of the market and socio-economic development, successful even if only with secondary data. The outcome of the study is a recommendation of why spatial autocorrelation models should be applied in market research. One should remember that they follow some challenges, since it is a time-consuming analysis (e.g. time needed to prepare a database with geographical attributes) with the lack of assurance to find a correlation and discover a new perspective on the results. Although some open source software is available, it is still demanding, whereas in Polish market agencies, SPSS is mainly used. One may ask if data visualization based only on geostatistics and ESDA is more comprehensible and appealing to the purchasers of the research studies.

However, the possible advantages are much more than that due to the demographic and geographic datasets available at a country-level. There is the possibility of an extra exploration of data and the possibility of finding a spatial pattern, spatial correlation that explains a bigger part of variance. That brings new conclusions based on the given data sets, as well as more advanced data visualization potential if the appropriate maps are available.

References

- Anselin, L., 1995, *Local Indicators of Spatial Association — LISA*, *Geographical Analysis*, 27(2), pp. 93-115, available at: www.drs.wisc.edu/people/faculty/curtis/documents/RS977/Anselin1995.pdf www.spatialanalysisonline.com/output/html/LocalindicatorsofspatialassociationLISA.html.
- Beck N., Gleditsch K., Beardsley K., 2006, *Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy*, *International Studies Quarterly*, 50, pp. 27-44.
- Bivand R.S., Edzer J.P., Gómez-Rubio V., 2008, *Applied Spatial Data with R*, Springer, New York; London.
- CSISS, 2015, Center for Spatially Integrated Social Science, available at: <http://www.csiss.org/GISPopSci/research/tools/spatial.php> (accessed May 5, 2015).
- Franzese R.J., Hays J.C., 2007, *Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data*, *Political Analysis*, 15(2), pp. 140-164.
- Franzese R.J., Hays J.C., Kachi A., 2012, *Modeling History Dependence in Network-Behavior Coevolution*, *Political Analysis*, 20(2), pp. 175-190.

- Jewczak M., Żółtaszek A., 2011, *Ekonometria przestrzenna i regionalne analizy ekonomiczne*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Miles S., 2012, *Spaces for Consumption*, London.
- Neumayer E., Plümper, T., 2013, *W*, EPSA 2013 Annual General Conference Paper 165, available at: <http://ssrn.com/abstract=2224687> (11.08.2014).
- Pietrzykowski R., 2011, *Wykorzystanie metod statystycznej analizy przestrzennej w badaniach ekonomicznych*, Roczniki ekonomiczne, 4, available at: http://www.kpsw.edu.pl/menu/pobierz/RE4/97_pietrzykowski.pdf.
- PMR, 2013, *Analiza potencjału handlowego średniej wielkości miast w Polsce 2013*, Kraków (document and database).
- Robinson A.C. et al., 2015, *Maps and the Geospatial Revolution: Teaching a Massive Open Online Course (MOOC) in Geography*, Journal of Geography in Higher Education, (April), pp. 1-18, available at: <http://www.tandfonline.com/doi/abs/10.1080/03098265.2014.996850>.
- Runge A., 2012, *Metodologiczne problemy badania miast średnich w Polsce*, Prace Geograficzne, 129, pp. 83-101.
- Schwartz D., 2015, *The Economics (and Nostalgia) of Dead Malls*, NY Times, available at: <http://www.nytimes.com/2015/01/04/business/the-economics-and-nostalgia-of-dead-malls.html>, (12.03.2015).
- Statystyczna analiza przestrzennego zróżnicowania rozwoju ekonomicznego i społecznego Polski*, 2013, eds. M. Trojak, T. Tokarski, 2013. WUJ, Kraków.
- Suchecky, B., 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Warszawa.
- Ward M.D., Gleditsch K.S., 2008, *Spatial Regression Models*, Sage, Thousand Oaks.