



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Detecting Local Variations in Spatial Interaction Models by Means of Geographically Weighted Regression

E. Nissi and A. Sarra

Dipartimento di Metodi Quantitativi e Teoria Economica, Viale Pindaro, 42 -65127, Pescara

Abstract: The development of local forms of spatial analysis has been the subject of intense research over last decade. The condition that observations are independent is not frequent in spatial data analysis. Parameter drift is often recognized when a model is calibrated separately across locations in space. In this study, we propose a local calibration procedure for handling varying parameters estimates of an origin-constrained spatial interaction model. In this context, the estimates of local parameters depend both on origins and destinations and a four dimensional space is involved. A suite of local parameters can be obtained by the maximisation of a weighted maximum likelihood function, exploiting the same principle of Geographically Weighted Regression (GWR) approach. Generally, the function for the weighting scheme in GWR uses only distance (geographical space) to determine weights. We propose a modified version of weighting function which takes into account both the spatial distance and a function of strength of connection between two specific destinations. The performance of different weighting functions for modelling the effects of spatial heterogeneity on flow data are compared and evaluated.

Key words: Spatial interaction models, geographically weighted regression, spatial non-stationarity, strength of connection function, migration flows

INTRODUCTION

Spatial interaction models focus on origin-destination pairs of regions and use flow data. They have been applied in many contexts in order to understand and explain movements of people, commodities, information, ideas, capital or knowledge from one set of places to another (migration studies, shopping, travel to work, airline passengers traffic). The simplest form of spatial interaction models is based on the analogy of Newton's law of gravity in physics. Basically, the flow between two places is a function of the ability of an origin to generate flows (the so-called propulsiveness), the capability of a destination to attract flows (the so-called attractiveness) and a sensible measure of separation of origin and destination (typically named spatial impedance). More solid foundations to spatial interaction models was provided by A.G. Wilson by their statistical mechanics entropy formulation, which ensures that these models can be cast in a probabilistic form (Roy, 2004). Traditionally, fitting the spatial interaction models to the observed data, is a question of estimating the unknown parameters: the ones characterising the propensity of each origin to generate flows, the ones characterising the attractiveness of each destination and the one related to distance deterrence effect. One way to enhance the spatial

interaction modelling is to properly take into account the parameters instability across the space, otherwise it would be possible to discover a severe misspecification bias in the general spatial interaction modelling formula (Tiefelsdorf, 2003). For example, if the distance decay parameter is not stable across the space, we can observe a counter-intuitive finding: an amplification of interaction as distance increases. Model parameters can be correctly interpreted only once spatial structure effects are under control. According to Fotheringham (1981), spatial structure can be defined as the size and configuration of the origins and destinations of regional system under investigation. The underlying spatial structure of spatial interaction models can be tackled from different points of view. Fotheringham (1983) proposes to introduce in the modelling a Competing Factor (CD) which represents the relationship between destination j and all other destinations, also named accessibility variable. The essential idea is that the movers are faced with a set of destinations competing for their attention: the more accessible a destination is to all other destinations in a spatial system, the less likely it is that destination is a terminating point for interaction from any given origin, *ceteris paribus*. The spatial proximity perspective expects competition only by spatial association. Lately, Jim (1998) proposes an alternative viewpoint on competition among

destinations, which contemplates the idea that destinations which are widely separated in space, or which vary greatly in size and/or distance from a given origin, may exert nevertheless identical influences on decision makers at that origin. Anyway, competing destinations models represent a good example for identifying of the misspecification global models and for improving the global statements of spatial behaviour. Another approach which attempts to accommodate spatial heterogeneity, by allowing the parameters of the model to vary with the spatial location of data, regards the expansion method of Casetti (1972) and Jones and Casetti (1992). Recently, Fischer and Griffith (2008) illustrate two approaches for a proper specification of spatial interaction models, facing with the problem of accounting for spatial autocorrelation effects. Firstly, following a spatial econometric perspective, they specify a spatial process for the disturbance terms reflecting origin and destination autoregressive dependence among origin-destination flows (Le Sage *et al.*, 2007). A second methodology relies on the eigenfunction spatial filtering approach developed by Griffith (2007). Chun (2008) shows the utility of eigenvector spatial filtering approach in a migration study. This study follows a GWR approach to detect spatial non stationarity of spatial interaction models. A modified equation for the computation of weights of GWR is also proposed. In particular, we recommend to incorporate in the weighting scheme both the spatial distance and a function of strength of connection between two specific locations. The potential advantages of our proposal are investigated by assessing the model's ability in replicating the known matrix flows.

SPATIAL NON-STATIONARITY AND LOCAL MODELS

The development of local forms of spatial analysis has recently gained in prominence (Anselin, 1995; Getis and Ord, 1996; Openshaw and Abrahart, 1996). The development of local models offers an excellent way to look at the local exception rather than broad generalisations, in order to understanding differences across space rather than similarities. In particular, local statistics allow relationships to be explored in more detail and will inevitably lead to a better understanding of spatial processes. As Tobler points out through the First Law of Geography: everything is related to everything else, but near things are more related than distant things. The quantitative outcome of this expression are the locational effects. They manifest themselves in two ways, namely, through spatial autocorrelation and spatial non-stationarity. Consequently, the observed values

drawn from spatial data sets not only will be spatially clustered (spatial autocorrelation), but also exhibit significant patterns of spatial non-stationarity, which is related to regional differentiations within the observational space: every location has an intrinsic degree of distinctiveness due to its situation with the respect to the rest of spatial system. Hence, spatial non-stationarity exists when the same stimulus provokes a different response in different part of the study region. Accordingly, the relationships can change across space, that is the relationship between variables under study is not constant from one location to the next. Very often, there are many unquantifiable factors which make one location different from another or involve extremely complex interactions that cannot be parsimoniously modelled. In literature, there are several approaches dealing with the problem of spatial non-stationarity. Among them, a possible technique to display spatial variations in relationships could be to parameterise a global model and look at the residuals to detect patterns. Hence, it is possible to use patterns in residuals to define patches and specify a separate equation for each patch or stratum. A different solution could be achieved allowing parameters to vary across space and this is the essence of Geographically Weighted Regression (GWR).

THE THEORETICAL FOUNDATION OF GEOGRAPHICALLY WEIGHTED REGRESSION

The GWR, developed by Brunsdon *et al.* (1996) and Fotheringham *et al.* (1998), is a non parametric methodology for the investigation of geographical drift of regression parameters. This technique extends the traditional regression model, by allowing the estimated coefficients to vary from location to location. The main contribution of the GWR methodology is the use of distance-weighted sub-samples of the data to produce locally linear regression estimates for every point in the space. Each set of parameter estimates is based on a distance-weighted sub-sample of nearby observations. The model has the general form:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad i = 1, \dots, n \quad (1)$$

where, (u_i, v_i) denotes the co-ordinates of i -th point in space and $\beta_k(u_i, v_i)$ is the local coefficient for the k -th explanatory variable at location i .

The location-specific regression coefficients $\beta_k(u_i, v_i)$ are functions of longitude and latitude coordinates u_i and v_i . More precisely $\beta_k(u_i, v_i)$ is a realisation of continuous function $\beta_k(u, v)$ at point i . The local parameters $\beta_k(u_i, v_i)$ are estimated by weighted least-squares procedure. The

weights w_{ij} are defined by a function of distance d_{ij} between a specific point j in the space at which data are observed and any point i in the space from which parameters are estimated.

The GWR approach gives more weights to data from observations close to i : data near to point i have more influence in the estimation of the $\beta_k(u_i, v_i)$'s than do the data located farther from i . Similar to kernel regression and kernel density estimation, the GWR estimates location-specific parameters using weighted least square estimation; the only exception regards the weights: they are based on locations in geographic rather than attribute space. A number, say n , of unknown vectors of local regression coefficients are to be estimated:

$$\beta_{(1)} = \begin{bmatrix} \beta_1(u_1, v_1) \\ \beta_2(u_1, v_1) \\ \vdots \\ \beta_k(u_1, v_1) \end{bmatrix}, \quad \beta_{(2)} = \begin{bmatrix} \beta_1(u_2, v_2) \\ \beta_2(u_2, v_2) \\ \vdots \\ \beta_k(u_2, v_2) \end{bmatrix}, \quad \beta_{(n)} = \begin{bmatrix} \beta_1(u_n, v_n) \\ \beta_2(u_n, v_n) \\ \vdots \\ \beta_k(u_n, v_n) \end{bmatrix}$$

The GWR estimates of the unknown local parameter vector $\beta_{(i)}$ are given by:

$$\hat{\beta}_{(i)} = [X'W_{(i)}X]^{-1} X'W_{(i)}y \quad i = 1, \dots, n \quad (2)$$

which provides estimates for each variable k and each geographical location i . In Eq. 2, $W_{(i)}$ is the $n \times n$ spatial weight matrix which has the form:

$$W_{(i)} = \begin{bmatrix} w_{i1} & \dots & 0 & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{bmatrix}$$

and whose diagonal elements are the weights of each observation. For example, w_{i1} is the weight of point 1 on the calibration of the model around point i . Thus the weighting of an observation is not longer constant in the calibration but varies with different locations. In the GWR framework, several weighting functions (kernels) can be considered and calibrated (Fotheringham *et al.*, 1998) for calculating the weighted least-squares estimators (2), although they tend to be Gaussian or Gaussian-like, reflecting the type of dependency found in most spatial processes. Whichever weighting function is used, the results will be sensitive to the degree of distance-decay (bandwidth). As the bandwidth increases, the parameter estimates tend to the estimates from a global model. In some parts of the region, where data are sparse, the local regressions may be based on relatively few data points and may not estimate the parameters reliably. Conversely,

fixed kernels in regions where data are dense may suffer from bias when the kernel are larger than needed. In order to avoid this problem, spatially adaptive weighting functions can be incorporated: these functions would have relatively small bandwidths in regions with high density of data points and larger in regions with low density of data points. Optimal bandwidth selection is a trade-off between bias and variance: too small a bandwidth leads to large variance in the local estimates; too large a bandwidth leads to large bias in the local estimates. Currently, Cross-Validation (CV) procedure, Generalised Cross-Validation (GCV) criterion of Loader (1999) or AIC (Akaike Information Criterion) (Hurvich *et al.*, 1998) are generally employed to select the finest value of bandwidth.

CALIBRATING A LOCAL SPATIAL INTERACTION MODEL BY MEANS OF THE GWR

From previous section, it is possible to argue that GWR provides a framework for evaluating how the strengths of relationships change with the spatial resolution of the analysis. Examples of application of GWR can be found in a variety of disciplines (health, social science, economy, urban economics), where it has been also applied as a graphical tool for data exploration. By contrast, there are few attempts to measure local variations in spatial interaction modelling (Nakaya, 2001). This neglect may be because spatial interaction models are more complex than models for the geographic distribution of attribute data, with each region being associated with several values as an origin as well as a destination. It is worth stressing that the spatial disaggregation of spatial flows constitutes one of the earliest examples of providing local information on relationships (Fotheringham and O'Kelly, 1989). In this study, we propose a local calibration procedure for handling varying parameter estimates of an origin-constrained spatial interaction model. Suppose we deal with a spatial system consisting of m origin regions and n destination regions. Let, y_{ij} denote observations on independent random variables, say y_{ij} , (where, i denotes the origin regions and j the destination regions) sampled from a specified probability distribution dependent upon some mean (today's prevailing specification is Poisson regression). The statistical spatial interaction model takes the general form as follows:

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} \quad (3)$$

where the mean μ_{ij} could be specified as a function of covariates measuring the characteristics of origin regions,

destination regions and their separation (Bailey and Gatrell, 1995). The origin-constrained model, which reflects destination effect and distance frictional effect, takes the general form as follows:

$$Y_{ij} = \alpha_i e^{v(x_i; \theta) + \gamma d_{ij}} + \varepsilon_{ij} \quad (4)$$

where, $v(x_i; \theta)$ is usually a linear function of the vector of destination characteristics (destination attractiveness); θ is a vector of associated parameters; the notation d_{ij} is used to represent the distance between i and j ; γ is a distance deterrence effect; α is the balancing factor to ensure the origin constraint on predicted flows. It is worth noting that in the spatial interaction context the estimates of local parameters depend both on origins and destinations. Hence, the understanding of spatial interaction local interactions can be difficult as a four dimensional space is involved: the geographical space in which flow origins and flow destinations are located. One way to derive a suite of local parameters could be to use the conventional approach of spatial interaction model separately for each origin in the spatial system of interest. This leads to recast the origin-constrained model outlined above as:

$$Y_{ij} = \alpha_i e^{v(x_i; \theta_{(i)}) + \gamma_{(i)} d_{ij}} + \varepsilon_{ij} \quad (5)$$

Local parameters can be obtained by the maximisation of a weighted maximum likelihood function, exploiting the same principle of Geographical Weighted Regression (GWR) approach, using an iterative algorithm, implemented with an ad hoc R routine. In Eq. 5 the parameters have the index of origin i , as they are calibrated using flows from each origin separately; the index between brackets indicates that the application of GWR principle refers only, for simplicity, to the destination locations. For each destination location, the log-likelihood for the model in Eq. 5 includes the geographical weights and it is specified as follows:

$$l_k = \left(\sum_{i=1}^m \left(Y_{ik} \left(\log \alpha_i + v(x_i; \theta_{(i)}) + \gamma_{(i)} d_{ik} \right) - \alpha_i e^{v(x_i; \theta_{(i)}) + \gamma_{(i)} d_{ik}} \right) w_{k(i)} \right) \quad (6)$$

As previously pointed out, the crucial issue regards the specification of the weighting function. The role of the weighting matrix in GWR is to represent the importance of individual observations among locations. Generally, the weight is determined by spatial distance only: it should decrease as difference between the focal point and its neighbours increases. One of the most commonly used spatial weighting function is the exponential distance-decay function:

$$w_{k(i)} = \exp(-d_{jk}^2 / h^2) \quad (7)$$

Function in Eq. 7 produces a decay of influence with distance. If j and k coincide, the weighting of data at that point will be the unity. The weighting of other data will decrease according to a Gaussian curve, as distance between j and k increases. This function has received many other specifications (Brunsdon *et al.*, 1996):

$$w_{k(i)} = \sqrt{\exp(d_{jk} / h)}$$

In this study we propose a modified version of weighting function which takes into account both the spatial distance and a function of strength of connection between two specific destinations. We assume that destinations which share more visitors tend to be more connected. Accordingly, we suggest the following format for the weighting function:

$$w_{k(i)} = \exp\left(-\frac{d_{jk}^2}{h^2} \times f(\text{strength of connection})\right) \quad (8)$$

The strength of connection between two specific destinations is defined by:

$$\frac{y_{kj}^2}{y_{ok} \times y_{oj}}$$

where, y_{kj} is the flow between k and destination j ; y_{ok} and y_{oj} denote the total flows of k and destination j respectively. We expect that the introduction in the weighting function of the strength of connection function can improve the model's ability to replicate the known matrix flow. In order to evaluate the effects of our approach on the estimates, the local spatial interaction models with spatial and spatial-strength of connection weighting functions are compared.

A review of spatial interaction literature suggest that it is possible to draw proper conclusion regarding model performance examining the values of goodness-of-fit measures. In this study, we make use of a combination of two statistics: the Standardised Root Mean Square Error (SRME) and the Information Gain (Knudnes and Fotheringham, 1986; Fotheringham and Knudnes, 1987). The first statistics is calculated as follows:

$$SRME = \frac{\left\{ \sum_i \sum_j (Y_{ij} - \hat{Y}_{ij})^2 / (m * n) \right\}^{\frac{1}{2}}}{\left(\sum_i \sum_j Y_{ij} / (m * n) \right)} \quad (9)$$

where, Y_{ij} is the observed number of people moving from region i to region j ; \hat{Y}_{ij} are the estimated interactions

between i and j and m and n are the number of origin and destination zones, respectively. The SRME statistic has a lower limit of zero, indicating perfectly accurate predictions and an upper limit variable, depending on the distributions of Y_{ij} 's, although, in practice, it is often 1. Instead, the information gain statistics is defined as:

$$IG = \sum_i \sum_j p_{ij} \ln(p_{ij}/q_{ij}) \tag{10}$$

where,

$$p_{ij} = Y_{ij} / \sum_i \sum_j Y_{ij}$$

and

$$q_{ij} = \hat{Y}_{ij} / \sum_i \sum_j \hat{Y}_{ij}$$

This goodness-of-fit statistic measures the difference between two probability distributions. It has minimum at zero, indicating a completely accurate set of predictions; the value of IG increases with differences between the observed and predicted flows distributions. In detecting local spatial variations of migration flows here, it is assumed that the spatial weighting functions, under study, are applied equally at each calibration point. The bandwidth, equal to 100, is determined on the basis of previous studies of phenomenon.

EXPLORING SPATIAL VARIATIONS IN MIGRATION FLOWS PATTERNS

Here, we employ a spatial interaction approach to relate flows of migrants among Polish provinces to some potentially important explanatory variables. Polish internal migration flows for the single year of 2004 and the geographic resolution of NUTS-2 (the 16 Polish provinces or voidvodships) constitute the empirical basis of our analysis. The data used in this study are drawn from the Polish Official Statistic (Polka Statystyka Publiczna). The aim of this application is twofold. Firstly, we estimate parameters of a global origin-constrained spatial interaction model in order to select variables which determine the relative attraction of destinations. Economic variables (such as GDP per capita), labour market variables (such as levels of employment), environmental variables, embracing physical, economic, social and political aspects, are seen as potentially important. Along with the so-called gravity variables, population and physical distance, also the unemployment rate and per capita GDP are incorporated in the fitted models as

covariates. Moreover a social explanatory variable, the rate of detectability of delinquents, is also added to the model to assess if better living conditions, in terms of social security, might have an impact on the migration process. Global research findings establish a substantial influence of economic factors on migration flows. The estimated coefficients for the unemployment rate and GDP per capita emphasise this aspect (Table 1, in which parameter coefficients are expressed in logarithms).

Secondly, a further purpose of this application concerns the formulation of two localised origin specific constrained models, to capture regionally different sensitivities of explanatory variables, exploiting the same principle of GWR. They rely on a different specification of weighting functions. In one of them, the weights are defined according to a common used Gaussian distance decay based function (Eq. 7); in the other, our proposed weighting function combining the geographical space and the strength of connection (Eq. 8) is adopted. The best way for understanding the geographical variations of parameters may be to build maps of local parameters. However, owing to the volume of output of local estimates, it is not possible to display all maps for each combination of origins and destinations. An informal but convenient indication of the extent of spatial non-stationarity in the local parameter estimates, has been obtained by assessing if the range of local estimates between the inter-quartile range is greater than that of 2 standard errors of the global estimates, as suggested in Fotheringham *et al.* (2002).

The main empirical findings, achieved by the calibration of the two localised origin-specific-constrained models, are summarised in Table 2, in which we set out for each sending region the median, the upper and lower quartiles and the minimum and maximum values of the estimated parameters, got from the estimates over all destination provinces. In Table 2, we do not display the distance decay exponent, whose impact is stable across the study region. The analysis of the overall results shows local anomalies of parameters drifts in some region, particularly in the area of capital Warszawa (Mazowieckie voidvodship). Particular attention is paid to the main economic determinants of migration flows: the unemployment rate and per capita GDP. Regarding that, it is interesting to point out that while the global coefficient

Table 1: Global origin-constrained model coefficients

Model	Value	SE	t-value	p-value
LOG POP	0.736	0.14203	5.185	0.000
LOG GDP	7.460	0.17836	41.828	0.000
LOG UNEMP	-4.860	0.04856	-100.102	0.000
LOG DELIQ	-4.754	0.04817	-98.695	0.000
LOG DIST	-0.002	0.00002	-123.615	0.000

Table 2: Summary statistics of estimated parameters for localised origin-specific–constrained models with different weighting functions

		POP		GDP		UNEMP		DELIQ	
		Strength-connection function	Exponential distance-decay function	Strength-connection function	Exponential distance-decay function	Strength-connection function	Exponential distance decay function	Strength-connection function	Exponential distance-decay function
Voidvodships (provinces)	Minimum	-10.160	-11.656	0.900	0.357	-11.390	-12.440	-10.910	-13.965
	Lwr Quartile	-8.230	-7.759	20.173	18.644	-10.305	-11.901	-9.970	-10.309
	Median	-7.505	-6.019	22.900	21.60	-10.165	-11.340	-9.630	-9.565
	Upr Quartile	-5.670	-3.85	23.553	24.647	-9.688	-10.999	-9.150	-9.223
	Maximum	3.680	4.856	25.400	28.527	-3.580	-3.288	-6.270	-5.437
Mazowieckie	Minimum	11.310	8.012	-33.940	-28.167	7.540	3.529	0.440	0.113
	Lwr Quartile	12.358	12.031	-22.463	-25.534	8.430	9.203	2.595	2.780
	Median	13.875	14.337	-20.085	-20.818	8.810	9.711	2.890	3.228
	Upr Quartile	15.133	17.528	-17.630	-17.552	10.485	10.628	3.363	3.702
	Maximum	23.900	18.897	-16.070	-9.262	13.000	12.122	3.940	7.843
Malopolskie	Minimum	-4.020	-6.331	-10.760	-14.168	-6.860	-7.737	-7.070	-9.130
	Lwr Quartile	3.888	6.556	-0.025	-3.004	-4.495	-2.923	-4.228	-3.376
	Median	4.890	6.849	1.550	-1.518	-4.155	-2.310	-4.045	-3.295
	Upr Quartile	6.178	8.117	2.945	-1.051	-3.973	-2.189	-3.538	-2.981
	Maximum	12.270	14.533	13.610	17.475	3.500	4.654	-0.840	2.652
Slaskie	Minimum	-8.840	-14.306	-3.790	-12.008	-4.690	-10.162	-4.230	-3.490
	Lwr Quartile	2.873	5.176	0.968	-1.174	-4.053	-4.990	-2.523	-1.054
	Median	3.075	5.432	2.655	-0.198	-3.825	-4.840	-2.385	-0.682
	Upr Quartile	4.325	6.821	2.965	0.070	-3.763	-4.766	-1.953	-0.567
	Maximum	8.730	15.605	17.140	23.703	-1.180	-2.694	-1.180	2.926
Lubelskie	Minimum	-9.630	-17.951	-25.790	-19.551	-18.740	-20.025	-12.150	-13.723
	Lwr Quartile	-8.108	-8.662	21.043	21.846	-14.595	-16.918	-9.653	-8.059
	Median	-7.695	-7.244	24.420	24.406	-14.075	-15.987	-9.215	-7.786
	Upr Quartile	-5.228	-5.092	24.978	25.742	-13.068	-14.798	-8.395	-7.156
	Maximum	25.120	15.964	28.500	37.986	-7.030	-2.943	9.450	2.635
Podkarpackie	Minimum	-0.480	-9.476	-14.370	-34.000	-10.840	-11.801	-10.740	-21.927
	Lwr Quartile	11.930	20.027	-12.565	-19.960	-7.050	-6.145	-8.385	-10.100
	Median	15.975	23.561	-9.045	-18.406	-6.370	-5.775	-7.550	-8.379
	Upr Quartile	19.210	25.978	-3.778	-12.566	-6.095	-4.783	-7.188	-7.476
	Maximum	21.880	46.324	14.570	28.074	-4.450	-3.227	-4.470	-5.777
Swietokrzyskie	Minimum	-17.930	-2.794	-20.250	-21.903	-14.240	-23.827	-10.080	-10.400
	Lwr Quartile	-3.045	4.536	-11.858	-13.555	-12.103	-13.412	-9.085	-8.416
	Median	6.725	13.210	-1.835	-2.795	-11.030	-12.768	-7.760	-7.390
	Upr Quartile	21.573	22.981	14.890	9.862	-9.663	-11.651	-6.980	-6.809
	Maximum	29.600	30.675	16.880	19.684	-8.820	-9.524	-4.050	-5.706
Podlaskie	Minimum	-20.560	-16.696	-1.480	-66.984	-16.750	-20.531	-16.030	-17.978
	Lwr Quartile	-5.953	-6.475	15.948	19.601	-5.050	-2.261	-10.393	-12.609
	Median	-5.255	-5.628	17.700	20.041	-4.690	-1.922	-10.155	-12.256
	Upr Quartile	-4.138	-5.392	19.160	21.077	-4.048	-1.254	-9.858	-11.873
	Maximum	-3.230	34.588	41.020	37.117	17.400	42.644	10.010	6.263
Wielkopolskie	Minimum	-16.990	-30.369	0.480	1.508	-3.880	-3.573	-13.850	-14.736
	Lwr Quartile	-13.315	-15.731	18.393	18.984	0.560	2.672	-8.925	-7.208
	Median	-10.850	-14.431	21.605	22.830	0.860	3.210	-8.250	-6.741
	Upr Quartile	-8.243	-11.196	24.705	25.018	1.148	3.556	-7.658	-6.429
	Maximum	14.310	1.253	28.540	45.315	10.060	12.648	-3.500	-2.689
Zachodniopomorskie	Minimum	-0.840	-10.496	-3.430	-36.157	-10.650	-14.705	-3.810	-2.682
	Lwr Quartile	5.743	0.012	-1.148	-0.560	-8.758	-11.820	-2.028	13.502
	Median	7.080	1.883	0.050	0.898	-8.340	-11.102	-1.750	14.172
	Upr Quartile	7.875	2.779	1.428	4.987	-8.108	-10.009	-1.495	14.383
	Maximum	10.390	32.444	8.890	15.596	-5.730	-4.783	1.030	15.805
Lubuskie	Minimum	-3.340	-11.015	-32.530	-40.504	-15.260	-22.631	-13.850	-21.750
	Lwr Quartile	20.225	-1.273	-26.838	-12.367	-6.835	-14.492	-9.618	-18.708
	Median	24.005	15.989	-18.190	4.514	-5.420	-11.970	-8.220	-15.832
	Upr Quartile	30.355	26.092	-13.158	20.469	-3.075	-8.375	-7.488	-13.169
	Maximum	39.300	45.436	25.710	47.769	5.530	-1.999	-3.600	-9.713
Dolnoslaskie	Minimum	-12.299	-10.754	7.063	-7.365	-9.665	-17.062	-3.565	-0.255
	Lwr Quartile	-7.778	-1.666	10.843	4.176	-7.823	-9.293	1.365	5.945
	Median	-6.484	-0.882	13.328	5.273	-7.060	-9.096	1.532	6.153
	Upr Quartile	-5.613	-0.127	15.945	6.605	-6.780	-8.449	2.075	6.396
	Maximum	-0.662	11.804	20.640	17.970	-2.139	-4.019	4.957	10.135
Opolskie	Minimum	-6.490	-1.891	-17.330	-26.401	-11.800	-6.378	-16.290	-15.986
	Lwr Quartile	15.973	18.849	-13.443	-21.323	-1.295	-3.379	-10.805	-12.213
	Median	18.045	22.714	-12.250	-16.958	-1.020	-2.297	-10.230	-11.004
	Upr Quartile	19.008	26.197	-8.645	-12.796	-0.870	0.017	-9.958	-9.684
	Maximum	22.850	30.298	28.550	11.479	2.030	3.545	-6.200	-4.486

Table 2: Continued

Voidvodships (provinces)		POP		GDP		UNEMP		DELIQ	
		Strength-connection function	Exponential distance-decay function	Strength-connection function	Exponential distance-decay function	Strength-connection function	Exponential distance decay function	Strength-connection function	Exponential distance-decay function
Kujawsko-Pomorskie	Minimum	-2.510	-8.202	0.530	-3.462	-9.880	-13.606	-6.130	-4.133
	Lwr Quartile	-0.268	1.087	5.323	-0.209	-8.275	-8.774	-4.610	-1.070
	Median	2.235	5.329	6.860	1.651	-7.890	-8.351	-3.655	-0.213
	Upr Quartile	3.673	6.887	10.145	7.605	-7.473	-7.882	-3.240	0.123
	Maximum	7.270	9.413	12.630	23.720	-5.710	-5.629	-2.340	0.710
Warminsko-Mazurskie	Minimum	-9.150	-17.136	10.320	15.177	-10.440	-12.412	-7.930	-9.324
	Lwr Quartile	-5.943	-11.739	14.448	17.326	-7.023	-5.886	-6.338	-6.653
	Median	-4.105	-8.218	15.070	19.978	-6.750	-5.699	-6.040	-5.457
	Upr Quartile	-3.443	-6.578	17.215	28.120	-6.478	-4.927	-5.825	-5.164
	Maximum	-1.570	-4.309	24.050	30.475	-2.920	-3.862	-3.190	1.176
Pomorskie	Minimum	-3.350	-4.538	-31.290	-40.797	-4.390	-1.563	-9.340	-5.655
	Lwr Quartile	-0.753	5.434	3.170	-7.763	-2.640	0.579	-4.728	-0.605
	Median	0.510	8.291	5.715	-5.740	-2.008	1.752	-4.090	0.270
	Upr Quartile	2.495	9.562	8.055	-2.455	-1.753	1.891	-3.798	0.778
	Maximum	20.040	25.202	10.170	9.865	14.770	18.910	2.710	6.434

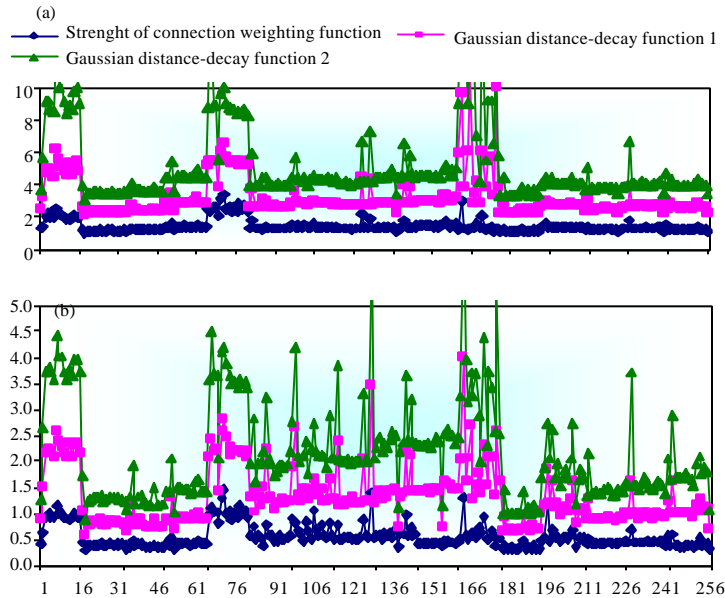


Fig. 1: Measures of predictive performance of localised specific origin-constrained models with different weighting functions

signs confirm that migrants are willing to move towards provinces characterised by low unemployment rate and high per capita GDP, the local estimates shows some counter-intuitive results. It should be noted that the geographical parameter drifts are strongly dependent on the combinations of origins and destinations. This makes very difficult to deeply recognize the determinants of people movements. Finally, to further illustrate the effect of our specification of the weighting function on the models ability in replicating the observed flow data, we calculate and compare two different goodness-of-fit statistics (SMRE and IG), as defined here. The potential advantages arising from the introduction of the function

of strength of connection in the weighting scheme, are evaluated by the comparison with the following different Gaussian weighting functions:

$$w_{k(i)} = \exp(-d_{jk}^2 / h^2)$$

and

$$w_{k(i)} = \sqrt{\exp(d_{jk} / h)}$$

indicated in Fig. 1 as Gaussian distance decay function 1 and Gaussian distance decay function 2, respectively.

The parameters of coefficients, estimated in each of the destination provinces and for each of the sending provinces allow us to obtain 256 matrices of predicted flows to be compared with the observed flows among Polish voivodeships. The results, obtained from the local spatial interaction models under study, indicate a better prediction power for the model with the new weighting function, through which we constantly observe smaller values of both the considered Standardised Root Mean Square Error and Information Gain statistics (Fig. 1). It implies that local spatial interaction model fits the data better if the spatial-strength of connection weighting function is used.

CONCLUSIONS

Although, applications of GWR can be found in a variety of disciplines, there are few efforts to document local variation of spatial interaction models. In this study, exploiting the same principle of GWR, we obtained the parameters drifts for localised origin-specific-constrained models. In the GWR approach, there are many options regarding the specification of the weighting function. Currently, the geographical weight designates the degree of spatial separation between the focal point and its neighbours. In this study, we recommend a diverse weighting function which includes both the geographical distance and the strength of connection between two regions, properly specified. Migration flows among Polish provinces are used as an illustrative application of the suggested approach. The obtained localised parameters exhibit a high degree of variability over space and demonstrate complex spatial patterns of the variables. We expected that the introduction in the weighting scheme of “the strength of connection” between two specific localisations could improve the prediction power of the model. In this regard, we compared the overall performance of our localised spatial interaction model, in terms of prediction errors, with localised spatial interaction models where the weighting functions are taken in the traditional Gaussian distance-decay form. The lower values of both goodness-of-fit statistics employed represent empirical evidence that the introduction of new weighting function has significantly improved the model’s ability in replicating the known matrix flows. In some way, the improvements of fitting can be regarded as an alternative way to detect the presence of spatial relationship, validating the usefulness of our proposed approach for handling varying parameters estimates of spatial interaction models.

REFERENCES

- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geograph. Anal.*, 27: 93-115.
- Bailey, T.C. and A.C. Gatrell, 1995. *Interactive Spatial Data Analysis*. Longman Ltd., Harlow, Essex, UK., pp: 413.
- Brunsdon, C.F., A.S. Fotheringham and M.E. Charlton, 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Anal.*, 28: 281-298.
- Casetti, E., 1972. Generating models by expansion method: Applications to geographical research. *Geographical Anal.*, 4: 81-91.
- Chun, Y., 2008. Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *J. Geographical Syst.*, 10: 317-344.
- Fischer, M.M. and D.A. Griffith, 2008. Modelling spatial autocorrelation in spatial interaction data: An application to patent citation data in the European union. *J. Regional Sci.*, 48: 969-998.
- Fotheringham, A.S., 1981. Spatial structure and the parameters of spatial interaction models. *Ann. Assoc. Am. Geographers*, 71: 425-436.
- Fotheringham, A.S., 1983. A new set of spatial interaction models: The theory of competing destinations. *Environ. Plann. A*, 15: 15-36.
- Fotheringham, A.S. and C.D. Knudnes, 1987. *Goodness-of-Fit Statistics. Concepts and Techniques in Modern Geography*. Geo Books, Norwich.
- Fotheringham, A.S. and M.E. O’Kelly, 1989. *Spatial Interaction Models: Formulations and Applications*. Kluwer Academic, Dordrecht.
- Fotheringham, A.S., M.E. Charlton and C.F. Brunsdon, 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environ. Plann. A*, 30: 1905-1927.
- Fotheringham, A.S., C. Brunsdon and M. Charlton, 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, Chichester, UK.
- Getis, A. and J.K. Ord, 1996. Local Spatial Statistics: An Overview. In: *Spatial Analysis: Modelling in a GIS Environment*, Longley, P. and M. Batty (Eds.). John Wiley, New York, pp: 261-277.
- Griffith, D., 2007. Spatial structure and spatial interaction: 25 years later. *Rev. Regional Stud.*, 37: 28-38.
- Hurvich, C.M., J.S. Simonoff and C.L. Tsai, 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Statistical Soc. Ser. B*, 60: 271-293.

- Jim, P., 1998. Competition among destinations in spatial interaction models: A new point of view. *Chinese Geographical Sci.*, 8: 212-224.
- Jones, J.P. and E. Casetti, 1992. Applications of the Expansion Methods. Routledge, London.
- Knudsen, C.D. and A.S. Fotheringham, 1986. Matrix comparison, goodness-of-fit and spatial interaction modeling. *Int. Regional Sci. Rev.*, 10: 127-147.
- Le Sage, J., M.M. Fischer and T. Scherngell, 2007. Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers Regional Sci.*, 86: 393-421.
- Loader, C., 1999. *Local Regression and Likelihood*. Springer, New York, pp: 308.
- Nakaya, T., 2001. Local spatial interaction modelling approach based on the geographically weighted regression approach. *GeoJ.*, 53: 347-358.
- Openshaw, S. and R.J. Abraham, 1996. Geocomputation. Proceedings of the 1st International Conference on GeoComputation, (GC'96), University of Leeds, Leeds, pp: 665-666.
- Roy, J.R., 2004. *Spatial Interaction Modelling: A Regional Science Context*. Springer-Verlag, Berlin, Heidelberg, New York.
- Tiefelsdorf, M., 2003. Misspecifications in interaction model distance decay relations: A spatial structure effect. *J. Geographical Syst.*, 5: 25-50.