Spatial structure effects in spatial interaction model: a Geographically Weighted Regression (GWR) approach

Eugenia Nissi
Dipartimento di Metodi Quantitativi e Teoria Economica
Viale Pindaro, 42 -65127 Pescara
nissi@dmqte.unich.it

Annalina Sarra
Dipartimento di Metodi Quantitativi e Teoria Economica
Viale Pindaro, 42 -65127 Pescara
asarra@dmqte.unich.it

Abstract: *The development of local forms of spatial analysis has been the subject of intense research over last decade. In this paper we propose a local calibration procedure for handling varying parameter estimates of an origin-constrained spatial interaction model. In this context, the estimates of local parameters depends both on origins and destinations and a four dimensional space is involved. A suitable estimation of local parameters can be obtained by the maximisation of a weighted maximum likelihood function, exploiting the same principle of geographical weighted regression (GWR) approach.*

Keywords: spatial interaction models, geographically weighted regression approach, spatial structure effects, migration flows, local estimates.

1.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").
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. 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. Interestingly, Fotheringham (1983) proposed to introduce in the modelling a competing factor (CD) which represents the relation between destination j and all other destinations, also named *accessibility variable*. Recently, Fischer *et al.*(2006) extended the traditional spatial interaction models to spatial econometric origin-destination flow models with an error structure to examine the role of spatial dependence in flows. This paper suggests an GWR approach to infer spatial nonstationarity of spatial interaction process. The remainder of the paper is

organized as follows. Section two, outlines details of geographical local modelling of flow data. In section three the results of an illustrative application are reported.

2.Local spatial interaction modelling: the GWR approach

The development of local forms of spatial analysis has been the subject of intense research over last decade. The geographically weighted regression (GWR), developed by Brundson, Fotheringham and Charlton (1996, 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 model has the general form

$$y_i = \beta_0(u_i v_i) + \sum_{i=1}^{p} \beta_k(u_i v_i)\, x_{ik} + \varepsilon_i \qquad i = 1,\ldots,n \tag{1}$$

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 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 to data located farther from *i*. In the GWR framework, different choices of spatial weighting function can be considered and calibrated (Fotheringham *et al*, 1996). Examples of application of GWR can be found in a variety of disciplines (health, social science, economy, urban economics). By contrast, there are few attempts to measure local variations in spatial interaction modelling (Nakaya, 2001). In this paper, 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). So, the statistical spatial interaction model takes the general form as follows:

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} \tag{2}$$

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 &Gatrell 1996).
The origin–constrained model, which reflects destination effect and distance frictional effect, takes the general form as follows:

$$Y_{ij} = \alpha_i e^{v(x_j;\theta)+\gamma d_{ij}} + \varepsilon_{ij} \tag{3}$$

where $v(x_j;\theta)$ 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 stressing that in the spatial interaction context the estimates of local parameters depend both on origins and destinations. 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 an estimation 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 :

$$Y_{ij} = \alpha_i e^{v(x;\theta_{i(j)})+\gamma_{i(j)}d_{ij}} + \varepsilon_{ij} \tag{4}$$

In the equation (4) 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 equation (4) includes the geographical weights and it is specified as follows:

$$l_k = \left(\sum_{i=1}^{m} \left(Y_{ik} \left(\log \alpha_i + v\left(x;\theta_{i(j)}\right) + \gamma_{i(j)} d_{ik} \right) - \alpha_i e^{v\left(x;\theta_{i(j)}+\gamma_{i(j)}d_{ik}\right)} \right) w_{k(j)} \right) \tag{5}$$

The crucial issue regards the specification of the weighting function. Generally, the weight is determined by spatial distance only: it should decrease as difference between the focal point and its neighbours increase. 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:

$$w_{k(j)} = \exp^{\left(-\frac{d_{jk}^{\;2}}{h^2} \times f\,(strenght\ of\ connection) \right)} \tag{6}$$

The $f\left(strenght\ of\ connection\right)$ is defined by $\dfrac{y_{kj}^2}{y_{ok} \times y_{oj}}$, 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. Another difficult regards how to choose an appropriate bandwidth in the weighted maximum likelihood equation (5). A kernel of radius h is positioned around the focal location k and the spatial interaction model fitted to the observations within distance h from k. Usually the most used approaches for selecting the "optimal kernel" are: a cross-validation procedure and the minimum AIC (Akaike Information Criterion) (Brunsdon, et al, 1998). It is also possible to predefine the bandwidth on the basis of the existing researcher's knowledge.

3.Exploring spatial variations in migration flows patterns: an illustrative application

In this section, the internal migration flows between 16 Polish regions over 2004 are used as illustrative example of local spatial interaction modelling. The original data are drawn from the Polish Official Statistics (Polka Statystyca Publiczna). "Pull factors" which might affect internal migration are firstly investigate by an origin-constrained model and then by its local specific version. A part from the so-called *gravity variables*, population and physical distance, the other covariates incorporated in the fitted model are mainly related to the labour market conditions: the unemployment rate and per capita GDP. Moreover a social covariate, 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. Our 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 aspects (see Table1, in which parameter coefficients are expressed in logarithmic). As known, the best way to interpret the geographical variations of parameters may be to build maps of local parameters. However, owing to the volume of output of these local estimates, we prefer here to summarise the main empirical findings of the calibration of a localised origin-specific–constrained model.

	Value	Std.Error	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 *1 Global origin-Constrained Model Coefficients*

The results indicate that the distance-decay parameter γ has almost the same elasticity throughout the study region. This means that a difference of distance from origins is not significant for migrants. A considerable number of studies argue that the impact of distance deterrence effect is negligible on internal migration flows. It is worth noting that the physical distance, used in this analysis, as a proxy to take into account costs related directly or indirectly to distance does not consider other important costs, such as time of moving, information costs and psychological costs. By contrast, the localised version of the spatial interaction model shows a distinctive parameter variations for the GDP and population coefficients. In particular, migrants from deprived areas prefer populous destinations, characterised by high performance of GDP per capita. Local anomalies of parameter drifts refer to the area of capital Warszawa. Less evident are the geographical patterns for the “attractiveness” expressed by the unemployment rate and the rate of detectability of delinquents. This study could be better addressed, by assessing to some extent the present specification of weighting function, which includes both the geographical distance and the strength of connection between two regions, and the selection of bandwidth (the same for each origin and equal to 100 here) might affect the results.

Bibliography

Bailey T.C., and Gatrell A.C. (1995) Interactive Spatial Data Analysis. Harlow, UK: Longman Scientific and Technical

Brunsdon C.F., Fotheringham A.S., Charlton M.E. (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. Geographical Analysis 28: 281-298

Fischer M.M., Griffith D.A, (2006) Modelling spatial autocorrelation in spatial interaction data. Available at SSRN:http://ssrn.com/abstract=1102183

Fotheringham A.S. (1981) Spatial structure and the parameters of spatial interaction models. Annals of the Association of American geographers 71 425-436

Fotheringham A.S. (1983) A new set of spatial interaction models: the theory of competing destinations. Environment and Planning A 15: 15-36

Fotheringham A.S., Charlton M.E., Brunsdon C.F. (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30: 1905-1927

Nakaya T. (2001) Local spatial interaction modelling approach based on the geographically weighted regression approach. GeoJournal 53: 347-358