Addressing the spatial nonstationarity in the MCI's parameters using Geographically Weighted Regression

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Departamento de Métodos Cuantitativos en Economía y Gestión

Suárez-Vega, Rafael Suárez-Vega[ab], José Luis Gutiérrez-Acuña[b], Manuel Rodríguez-Díaz[bc]
[a]Departamento de Métodos Cuantitativos en Economía y Gestión
[b]Instituto Universitario de Turismo y Desarrollo Económico Sostenible (Tides)
[c]Departamento de Economía y Dirección de Empresas
Universidad de Las Palmas de Gran Canaria

Abstract

In this paper, a Multiplicative Competitive Interactive Model (MCI) is treated. Although several papers have addressed this type of model, including different ways for estimating the parameters involved, a very few documents have faced the spatial nonstationarity problem that may arise. Nakanishi and Cooper (1974) proposed one of the most used procedures for estimating the MCI parameters. They suggested some transformation in order to estimate the parameters by means ordinary least squares (OLS). This method does not take into account the possible variation of the parameters because they are constant along the working space. Therefore, in order to consider the spatial nonstationarity we propose the use Geographically Weighted Regression (a local lineal regression method) instead the global ordinary least square. An application revealing the spatial nonstationarity of the OLS parameters is presented. This example is also used to show how the local regression model predicts data better than the OLS model.

1. MCI models:

Different location models and procedures for calculating trading areas have been proposed in the specialized literature. The gravity models has widely used within the field of the retail distribution. These models are based on the assumption that individual movements between points are inversely proportional to the distance between them. Multiplicative Interactive Models (MCI) are a generalization of the gravity models to the case where the attraction perceived by the customers from the facilities depends on other variable in addition to the distance. For MCI models, the probability that a customer at i buys at a facility j is given by

$$p_{ij} = \frac{U_{ij}}{\sum_{j=1}^{n} U_{ij}},$$

$$U_{ij} = \prod_{k=1}^{K} \left(X_{kj}^{\alpha_k} \right) d_{ij}^{\lambda},$$

where U_{ij} is the utility of facility j to customers at point i, X_{kj} is the kth of K variables measuring characteristics of facility j, α_k are the parameters reflecting the effect of kth store characteristics, and λ is the parameter reflecting the transportation cost effect.

Nakanishi and Cooper (1974) proposed the following log-transformed-centered form to obtain least squares estimates of the parameters:

$$\ln\left(p_{ij} / \tilde{p}_{i\cdot}\right) = \sum_{k=1}^{K} \alpha_k \ln\left(X_{kj} / \tilde{X}_{k\cdot}\right) + \lambda \ln\left(d_{ij} / \tilde{d}_{i\cdot}\right)$$

where $\tilde{p}_{i\cdot}$, $\tilde{X}_{k\cdot}$, $\tilde{d}_{i\cdot}$ are the geometric means of p_{ij}, X_{kj}, and d_{ij} over j, respectively.

These parameters are influenced by socio-demographic characteristics of the market. Local differences in this aspect (changes in density population or income ratios) may produce local differences in customers' perception of the facilities attraction. (See Gosh (1984) for a deeper discussion about nonestationarity in MCI models)

2. Using GWR for estimating MCI parameters

GWR (Brunsdon et al. (1996) and Fotheringham et al. (1996, 1997)) is based on the assumption that closer customers present closer preferences. Following this method, parameters for a particular location must be estimated under the assumption that nearby observed data are more influent than those located further away. So the traditional OLS can be rewritten as

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

were (u_i,v_i) represents the coordinates of the ith point in space and $\beta_k(u_i, v_i)$ is a realization of the continuous function at point i.

Let X be the matrix of explanatory variables (including the constant) and Y the vector of dependent variable, the estimation of the vector β of coefficients regressors is

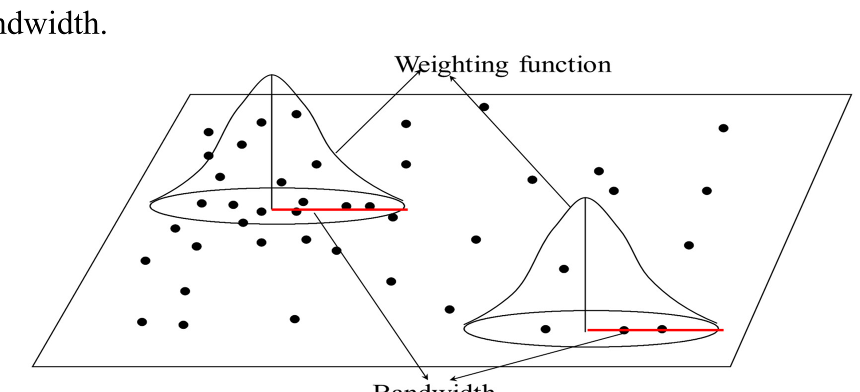
$$\hat{\beta}(u_i, v_i) = \left(X^T W(u_i, v_i) X \right)^{-1} X^T W(u_i, v_i) Y,$$

with $W(u_i,v_i)$ a weight matrix where each element w_{jj} in the diagonal represents the weight of the observation j (j = 1, 2,…, n) for estimating the parameters in location (u_i,v_i).

The function used for obtaining $W(u_i,v_i)$ is named kernel. In this application a bi-square (kernel) was selected

$$w_{jj} = w_j(u_i, v_i) = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2, & \text{if } d_{ij} < h \\ 0, & \text{if } d_{ij} > h \end{cases},$$

where h is the bandwidth.



(Imagen taken from "Geographically weighted regression", Yu and Wei (2004)

3. Particular case: Huff model

If spatial variation for the parameters model in Huff (Huff (1964)) is taken into account, probabilities p_{ij} can be reformulated as:

$$p_{ij} = \frac{\dfrac{S_j^{\alpha_i}}{d_{ij}^{\lambda_i}}}{\sum_{j=1}^{n} \dfrac{S_k^{\alpha_i}}{d_{kj}^{\lambda_i}}}$$

where α_i and λ_i are the estimated parameters reflecting the effect of the sales surface and the distance for customers located at demand node i.

Steps given for estimating α_i and λ_i :

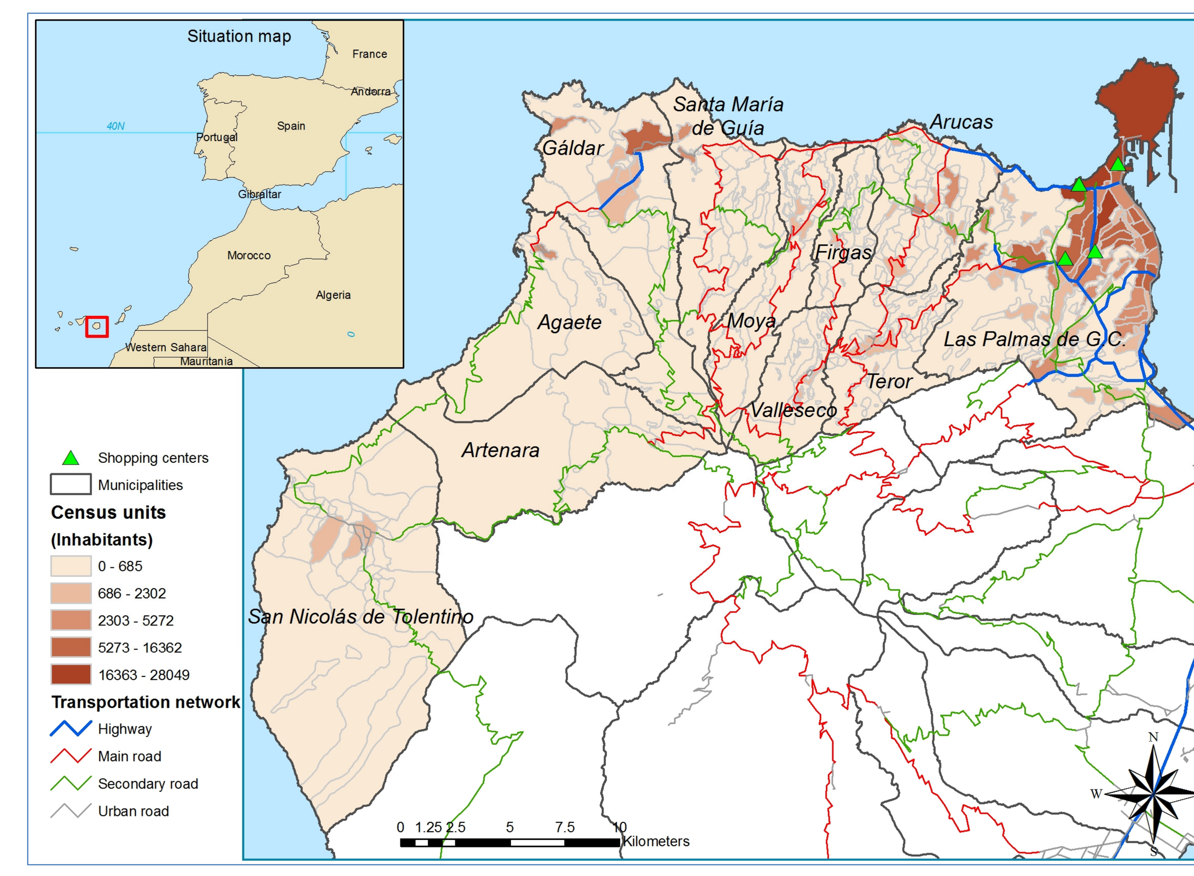
- Obtain revealed probabilities p_{ij} from a survey.
- Select the kernel and the bandwidth for the GWR model that better fits the sample data.
- Estimate using this settlement parameters α_i and λ_i for the demand point.

Given the parameters, the probabilities that customers at i purchase at a new facility with sales surface S and located at P is given by:

$$p_{iP} = \frac{\dfrac{S^{\alpha_i}}{d_{iP}^{\lambda_i}}}{\dfrac{S^{\alpha_i}}{d_{iP}^{\lambda_i}} + \sum_{j=1}^{n} \dfrac{S_k^{\alpha_i}}{d_{ij}^{\lambda_i}}}$$

Therefore, if w_i is the buying power at demand node i, the estimated capture for the new facility is

$$MS(P) = \sum_{j=1}^{n} w_i p_{iP}$$



Situation map and study area

4. Application

In this study we have analyzed the potential market share of a new grocery store considering the market formed by the north part of the island, the capital (Las Palmas de Gran Canaria) and nine municipalities which form the Northwest Community. This study area presents the particularity that there exists a significant socio-demographic difference between the capital of the island, an urban municipality with a population of 381114 inhabitants, and the rest of municipalities which are mainly rural areas with a total population of 119536.

The figure above shows the population distribution by census units for the study area and the shopping centers considered. The population of each census unit was allocated to the gravity center of the housing units sited in it.

Grocery stores, shopping centers, population gravity centers and polls were georreferencied in order to calculate transportations costs and estimated the GWR parameters.

5. Results

For calibrating the Huff model, a survey formed by 726 valid questionnaires was used. People were asked about the proportion of food purchase they made in the five shopping centers considered in the study. The remaining purchase power was considered as proximity purchase. The proximity purchase was allocated to the closest grocery store.

GWR over sample data
(A bi-square kernel with a 12000 m. fixed bandwidth was selected for being the settlement that better fits the sample data)

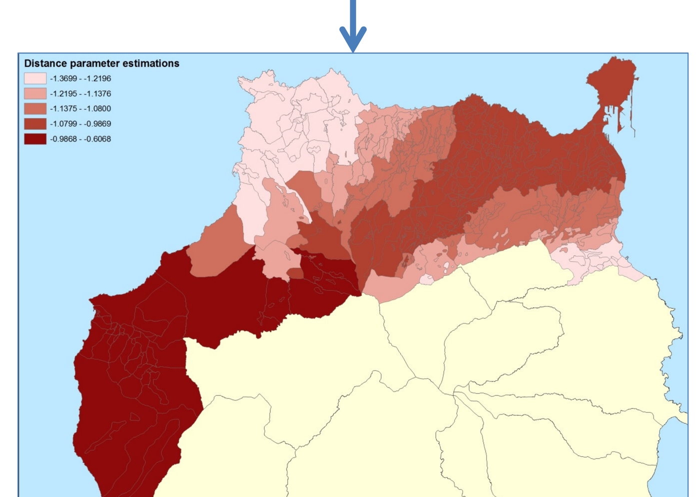
	OLS		GWR				
	Estimated	p-value	Average	Min	Max	STD	p-value
α	-0.573	0.0000	-0.5387	-0.9675	-0.3410	0.1246	0.0018
λ	-1.083	0.0000	-1.1067	-1.3180	-0.8155	0.1008	0.0168
AICc	18792		18770				
	0.3859		0.3898				
Improvement F-test p-value: 0.018							

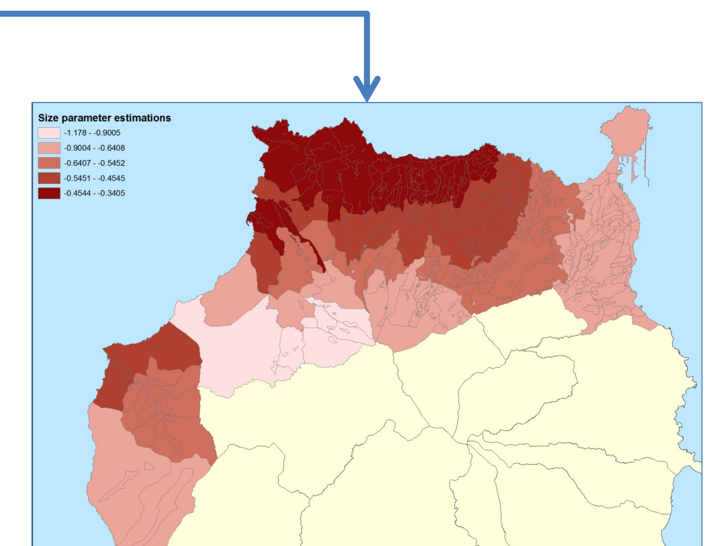
Results for OLS and GWR estimations for the sample data

Spatial nonstationarity exists for the two parameters (F-test proposed by Leung et al. (2000)). It is most significant for the parameter associated to the size.

GWR model better fits data sample than OLS model (F-test proposed by Fotheringham et al. (2002))

GWR over demand points



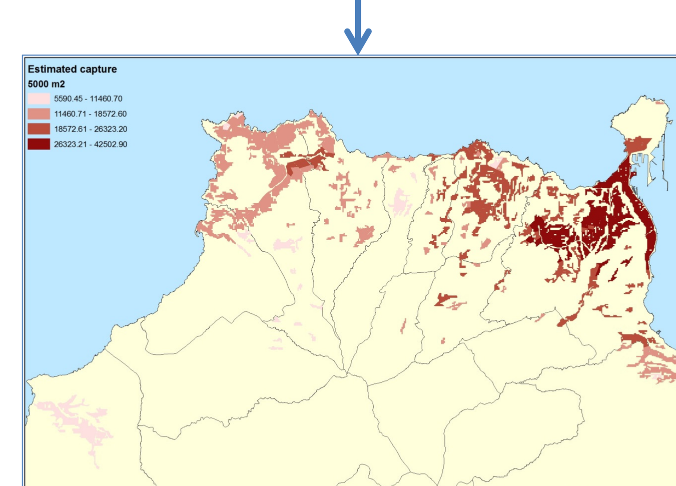


Estimated parameters for the demand node

	Average est.	Min.	Max.	STD
α	-0.5617	-1.1781	-0.3405	0.1440
λ	-1.0945	-1.3699	-0.6068	0.1040

GWR parameters estimations for the demand points

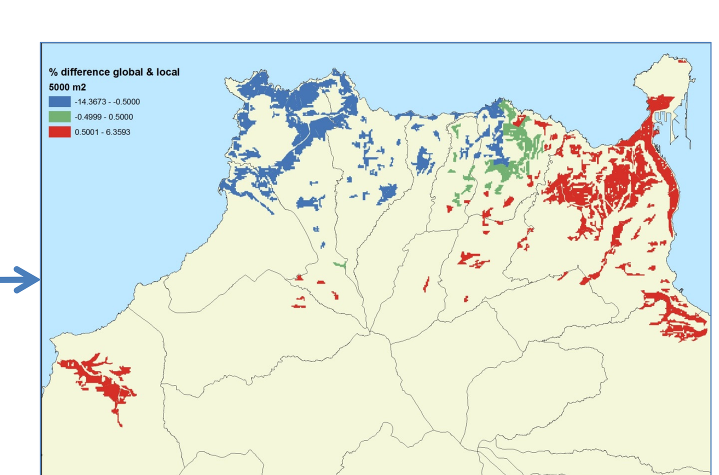
These estimations were obtained using the kernel and bandwidth selected for estimating the sample data.



Local estimated capture for a new 5000 m² grocery store

Using the estimated capture map not only the site with the maximum capture can be identified but also an overall vision of the profitability of the different zones in the feasible area is obtained.

$$100 * \frac{(GC - LC)}{GC}$$



Percentage difference between global capture (*GC*) and local capture (*LC*)

The map shows as the OLS model overestimates the capture for the new store in the most populated area and in the most remote zone (the western part of the feasible region). Nevertheless, in the more accessible zone out of the capital, the GWR model obtains higher capture estimations than OLS model.

6. References

- Brunsdon C, Fotheringham AS, Charlton M (1996): Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis* 28(4):281-298.
- Fotheringham AS, Brunsdon C, Charlton M (1996): The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems* 10:605-627.
- Fotheringham AS, Charlton M, Brunsdon C (1997): Two techniques for exploring nonstationarity in geographical data. *Geographical Systems* 4:59-82.
- Ghosh, A. (1984): "Parameter nonstationarity in retail choice models", *Journal of Business Research* 12: 425-436.
- Huff DL (1964): "Defining and estimating a trading area", Journal of Marketing 28: 34-38.
- Leung, Y., Mey, C.L., Zanhg, W.X. (2000): "Statistical test for special nonstationarity based on geographically weighted regression model*", Environmental Planning A* 32: 9-32.
- Nakanishi M and Cooper LG (1974): "Parameter estimation for a Multiplicative Competitive Interaction Model-Least Square Approach", *Journal of Marketing Research* 11, 303-311.

RED LOCA
RED DE LOCALIZACIÓN Y PROBLEMAS AFINES
Workshop on Locational Analysis and Related Problems 2013
X Anniversary of the Spanish Location Network
Torremolinos (Málaga), June 19-21, 2013