*Environment and Planning A* 2002, volume 34, pages 883 – 904

DOI:10.1068/a34133

# A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests

**Antonio Páez**
Center for Northeast Asian Studies, Tohoku University, Kawauchi, Aoba-ku, Sendai 980-8576, Japan; e-mail: antonio@rs.civil.tohoku.ac.jp

**Takashi Uchida**
Graduate School of Engineering, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 9558-8585, Japan; e-mail: uchida@civil.eng.osaka-cu.ac.jp

**Kazuaki Miyamoto**
Center for Northeast Asian Studies, Tohoku University, Kawauchi, Aoba-ku, Sendai 980-8576, Japan; e-mail: miyamoto@plan.civil.tohoku.ac.jp
Received 14 June 2001; in revised form 18 December 2001

**Abstract. Spatial association effects, perhaps the most important concern in the analysis of spatial data, have been amply studied from a global perspective in the exploratory and modeling domains, and more recently also from a local perspective in the realm of exploratory data analysis. In a local modeling framework, however, the issue of how to detect and model spatial association by using geographically weighted regression (GWR) remains largely unresolved. In this paper we exploit a recent development that casts GWR as a model of locational heterogeneity, to formulate a general model of spatial effects that includes as special cases GWR with a spatially lagged objective variable and GWR with spatial error autocorrelation. The approach also permits the derivation of formal tests against several forms of model misspecification, including locational heterogeneity in global models, and spatial error autocorrelation in GWR models. Application of these results is exemplified with a case study.**

## 1 Introduction

Spatial association effects, mainly under the guise of spatial patterns and error autocorrelation, but also as spatially autoregressive processes, are one of the most important concerns in the analysis of spatial data. From a global perspective, the study of spatial association has a long and reputable history of development in the exploratory (Moran, 1948; Geary, 1954; Cliff and Ord, 1973; 1981) as well as in the modeling domains (Cliff and Ord, 1981; Anselin, 1988a; Griffith, 1988; Haining, 1990). Following a recent trend in spatial analysis that has emphasized the local as opposite to the global, there have also been important breakthroughs in the past few years in the analysis of local spatial association through the use of contiguity-block statistics (Nass and Garfinkle, 1992), distance statistics (Getis and Ord, 1992; Ord and Getis, 1995; 2001), and the local decomposition of global statistics of spatial association (Anselin, 1995).

In the spatial local modeling domain, on the other hand, the issue of how to detect and model spatial association by using the technique of geographically weighted regression (GWR) remains largely unresolved. Brunsdon et al (1998), for instance, proposed a spatially autoregressive model that incorporates the concept of geographical weights, but concentrated mainly on some important estimation issues while stopping short of presenting the results needed to conduct inference of the model. Leung et al (2000), on the other hand, developed a method to detect spatial autocorrelation among the residuals of GWR. We will argue, however, that the nature of their test is such that one cannot be sure whether the effect, if and when detected, corresponds to genuine spatial error autocorrelation, an omitted spatially lagged objective variable, or heterogeneity.

It is our objective to address this perceived gap in the literature of GWR, and local spatial analysis in general, and to provide a general framework to estimate, test, and diagnose a GWR model that includes spatial association components. In a previous paper (Páez et al, 2002), we introduced a variance function based on the concept of geographical weights developed by Fotheringham, Brunsdon, and Charlton (Brunsdon et al, 1996; Fotheringham et al, 1997; 1998), and developed a method to model locational heterogeneity (which implies parametric nonstationarity) and to obtain locally linear parameter estimates. In addition, we also derived a statistical test for locational heterogeneity, and showed it to be in fact a test for heteroscedasticity. In order to focus on the conceptual aspects of the method, however, other methodological issues regarding spatial association were not covered there. In this paper we extend those previous results and formulate a general model of spatial effects that includes as special cases GWR with a spatially lagged objective variable (GWR-SL) and GWR with spatial error autocorrelation (GWR-SEA). It will be seen that the approach leads to formal tests against several forms of model misspecification that include locational heterogeneity in global models, and spatial error autocorrelation in GWR models.

The outline of the paper is as follows. In the following section we describe a general model for spatial processes, and show how GWR is a particular case of a classical spatial specification. In section 3 we discuss estimation of two important model forms, namely GWR-SL and GWR-SEA. In section 4 we introduce the elements of the asymptotic variance matrix required to conduct inference and to test the specification of the model. In section 5 we derive a wide array of model specification tests, to test for locational heterogeneity in global models, and to test for an omitted spatial lag and spatial error autocorrelation in GWR models. Finally, in section 6, we apply the models and tests derived in this paper to a case study, in order to show how they can be fruitfully used to conduct local spatial analysis and modeling.

## 2 A general spatial model

The linear GWR formulation (Brunsdon et al, 1996; Páez et al, 2002), and the GWR-SL formulation (Brunsdon et al, 1998; Páez et al, 1999) so far developed in the literature can be considered as two particular cases of a general model of spatial processes for cross-sectional data. To see this, we borrow from Anselin's notation (1988a) to define a general spatial model as follows:

$$\boldsymbol{Y} = \rho \mathbf{W}_1 \boldsymbol{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{1}$$

$$\boldsymbol{\varepsilon} = \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \boldsymbol{\mu}\,, \tag{2}$$

with

$$\boldsymbol{\mu} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Omega})\,. \tag{3}$$

In this model, $\boldsymbol{Y}$ is a vector $(n \times 1)$ of objective variable observations, $\mathbf{X}$ is a matrix $(n \times K)$ of explanatory variables that include the usual constant term, and $\boldsymbol{\beta}$ is a vector $(K \times 1)$ of parameters corresponding to $K$ explanatory variables. Scalars $\rho$ and $\lambda$ are parameters of spatial association corresponding to a spatially lagged objective variable and spatial error autocorrelation, in that order. Vectors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are spatially autocorrelated and independent error terms, respectively, with the independent error terms defined by the normal distribution with mean 0 and a general covariance structure given by matrix $\boldsymbol{\Omega}$ (diagonal with elements $\omega_{ii}$). A common simplification is to assume a constant variance ($\boldsymbol{\Omega} = \sigma^2\mathbf{I}$); heterogeneity, however, can be implemented if nonconstant variance functions are introduced, as explained below. Spatial interaction matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ $(n \times n)$ in the above formulation are exogenously defined on the basis of the geographical configuration of the observations and some interaction criterion (that is, $\mathbf{W}_1$ and $\mathbf{W}_2$ are nonstochastic)). Moreover, they are usually row standardised so that the sum of all values in a row is 1 and each value represents a proportion of the row total. This is done for ease of interpretation on the one hand, and for technical reasons regarding the regularity conditions of the likelihood function on the other, as discussed below. The elements of the matrix are positive in the entries corresponding to locations that interact and 0s in the diagonal and elsewhere (for technical details see Cliff and Ord, 1981). Two distinct spatial interaction matrices are introduced in order to allow for the possibility that each process is driven by a different class of interactions. For example, one matrix could be defined on a first-order criterion (nearest neighbor) and the other on a second-order criterion, to test different theories of interaction; or one could be defined by using a lineal distance-decay criterion and the other quadratic distance decay. A simpler case results from letting $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$. The above general model potentially covers three different spatial processes.

1. A spatially lagged objective variable (or substantive spatial association), represented by the term associated with parameter $\rho$. This form of spatial association can be interpreted as a type of economic (Murdoch and Sandler, 1997; Murdoch et al, 1997) or spatial externality (Griffith, 1999), as information spillovers (Fingleton, 2000) or as an index of information content (Griffith and Layne, 1999).
2. Spatial error autocorrelation, represented by the term associated with $\lambda$ in the definition of the error terms. The interpretation of association of this form is usually as a surrogate for unobservable variables that follow some meaningful spatial pattern.
3. Spatial heterogeneity, modeled through the covariance structure of the error terms. This effect is interpreted as structural instability (that is, nonstationarity) of the process, or in other words, as regional or local variability in the operation of the process, which can be modeled in a discontinuous fashion (Páez et al, 2001), or in a continuous fashion using an absolute or relative reference framework. We term the latter locational heterogeneity (Páez et al, 2002).

If we introduce the following definitions:

$$\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1 , \tag{4}$$

$$\mathbf{B} = \mathbf{I} - \lambda \mathbf{W}_2 , \tag{5}$$

then the log-likelihood function for model (1)–(3) can be shown to be (Anselin, 1988a; pages 61–63):

$$L = -\frac{n}{2}\ln \pi - \frac{1}{2}\ln|\mathbf{\Omega}| + \ln|\mathbf{B}| + \ln|\mathbf{A}| - \frac{1}{2}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} , \tag{6}$$

with

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = (\mathbf{A}\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{\Omega}^{-1}\mathbf{B}(\mathbf{A}\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}) , \tag{7}$$

as the sum of the appropriately transformed squared errors. The above function leads to a well-behaved optimization problem, although the presence of determinants $|\mathbf{\Omega}|$, $|\mathbf{B}|$, and $|\mathbf{A}|$ can cause some problems, with the underlying issue being that asymptotic properties for the estimates will only hold if the regularity conditions for function (6) are satisfied. Fortunately, regularity can be attained under fairly general conditions; in the context of spatial models, Anselin (1988a, page 63) notes that it can be achieved if the following partial constraints are satisfied:

$$|\mathbf{I} - \rho \mathbf{W}_1| > 0 , \tag{8}$$

and

$$|\mathbf{I} - \lambda \mathbf{W}_2| > 0\,, \tag{9}$$

$$\omega_{ii} > 0, \qquad \forall i\,. \tag{10}$$

The use of row-standardized $\mathbf{W}_\#$ matrices, a common situation in practical applications, combined with a restriction of the spatial parameters to the ranges $1/w_1^{\min} < \rho < 1$, and $1/w_2^{\min} < \lambda < 1$, is usually enough to satisfy the first two constraints ($w_\#^{\min}$ is the minimum eigenvalue of matrix $\mathbf{W}_\#$). The third constraint, on the other hand, is the usual nonnegativity condition of the variance.

Following the same line of reasoning as in Páez et al (2002), we now propose a form of the covariance matrix as follows:

$$\mathbf{\Omega} = \sigma^2 \mathbf{G}\,, \tag{11}$$

with elements as:

$$\omega_{ii} = \sigma^2 \mathrm{g}_i(\gamma, z_i), \quad \text{and } \omega_{ij} = 0, \qquad \forall i, j, \quad i \neq j\,. \tag{12}$$

This formulation posits a model for the variance that is a function of a $p \times 1$ vector of known variables $z_i$, an unobservable parameter vector $\gamma\,(p \times 1)$,. and an unknown constant $\sigma^2$. The geographically weighted specification is obtained by defining a variance model of the exponential form as follows:

$$\mathrm{g}_{oi}(\gamma_o, a_{oi}) = \exp(\gamma_o d_{oi}^2)\,. \tag{13}$$

In this case $p = 1$, and known variable $d_{oi}$ is the distance between point $o$ (hereafter the focal point, which can be any location in the study area) and observation $i$ ($i = 1, \ldots, n$). The focal point is the origin of a relative reference framework, and distance to the obervations is given from this origin. Parameter $\gamma_o$ is called the kernel bandwidth after the fashion of GWR terminology, noting that function (13) is the inverse of the kernel function of conventional GWR modeling (see Brunsdon et al, 1996).

Three important characteristics of function (13) are worth noting. The first and most important for the case at hand is that the proposed model for the variance satisfies the nonnegativity constraint given by equation (10) as long as variance parameter $\sigma^2$ is positive. This means that regularity conditions for the likelihood function will hold whenever constraints (8) and (9) are also at least partially satisfied. In this sense, exponential function (13) is only one among a number of possible alternatives, and there are a number of valid functions that satisfy the nonnegativity constraint, most importantly those used for variogram modeling in geostatistics (Chilès and Delfinier, 1999). A second characteristic of function (13) is that the alternative model of constant variance ($\mathbf{\Omega} = \sigma^2\mathbf{I}$) results when $\gamma_o = \gamma_o^* = 0$, which means that the hypothesis of homogeneity (which implies spatial stationarity) can be easily defined within a testing framework. Third, the variance is not a function of absolute, but of relative location with respect to focal point $o$, and thus describes what we have termed locational heterogeneity. By convention, parameters estimated in this fashion are assigned to corresponding location $o$, thus resulting in locally linear parameter estimates, as opposed to global parameter estimates (see Brunsdon et al, 1996). Locational subindex $o$ ($o = 1, \ldots, m$) is adopted to indicate that parameters are location specific ($m$ is the number of local models being estimated).

Exponential variance functions such as (13) have been studied by, among others, Box and Meyer (1986), Verbyla (1993), and Lyon and Tsai (1996). The difference with these studies is that here the variance is defined as a function of spatial ordering, but more importantly, that the reference taken for ordering the observations is relative to a focal point selected by the analyst. In this way, local variability of the process can be detected and modeled using the framework presented in Páez et al (2002) and in the present paper.

Now a general GWR model can be defined in terms of local parameters as follows:

$$\boldsymbol{Y} = \rho_o \mathbf{W}_1 \boldsymbol{Y} + \mathbf{X}\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o \,, \tag{14}$$

$$\boldsymbol{\varepsilon}_o = \lambda_o \mathbf{W}_2 \boldsymbol{\varepsilon}_o + \boldsymbol{\mu}_o \,, \tag{15}$$

or alternatively, if we define $\mathbf{A}_o = \mathbf{I} - \rho_o \mathbf{W}_1$, and $\mathbf{B}_o = \mathbf{I} - \lambda_o \mathbf{W}_2$, as:

$$\mathbf{A}_o \boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o \,, \tag{16}$$

$$\mathbf{B}_o \boldsymbol{\varepsilon}_o = \boldsymbol{\mu}_o \,. \tag{17}$$

Matrices $\mathbf{A}_o$ and $\mathbf{B}_o$ are now subindexed because they depend on local parameters $\rho_o$ and $\lambda_o$, respectively. Now, the covariance structure of the error terms $\mu_o$ is given by:

$$\boldsymbol{\Omega}_o = \sigma_o^2 \mathbf{G}_o \,, \tag{18}$$

where $\mathbf{G}_o$ is a diagonal matrix with elements given by function (13), which depends in turn on parameter $\gamma_o$. It is simple to show that the log-likelihood function for this model is:

$$L_o = -\frac{n}{2}\ln \pi - \frac{n}{2}\ln \sigma_o^2 - \frac{1}{2}\ln |\mathbf{G}_o| + \ln |\mathbf{B}_o| + \ln |\mathbf{A}_o| - \frac{1}{2}\boldsymbol{v}_o^{\mathrm{T}}\boldsymbol{v}_o \,, \tag{19}$$

with:

$$\boldsymbol{v}_o^{\mathrm{T}}\boldsymbol{v}_o = \frac{1}{\sigma_o^2}(\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{B}_o^{\mathrm{T}} \mathbf{G}_o^{-1} \mathbf{B}_o (\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o) \,. \tag{20}$$

The model has a total of $4 + K$ parameters (matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are nonstochastic), which can be expressed in vector form as $\boldsymbol{\theta}_o = [\rho_o, \beta_o^{\mathrm{T}}, \lambda_o, \sigma_o^2, \gamma_o]^{\mathrm{T}}$. Function (19) is clearly a particular case of the more general function (6) that results from adopting a variance function based on the concept of geographical weights. The above function is the basis for the estimation and inference procedures discussed in the following sections.

## 3 Estimation of models of spatial effects

### 3.1 First-order conditions

In order to obtain expressions for the estimators, we need to obtain the first-order conditions of the log-likelihood function above (that is, the conditions under which the log-likelihood function is maximized). This is done by taking partial derivatives of function (19) with respect to the elements in vector $\boldsymbol{\theta}_o = [\rho_o, \boldsymbol{\beta}_o^{\mathrm{T}}, \lambda_o, \sigma_o^2, \gamma_o]^{\mathrm{T}}$. The resulting vector of partial derivatives, or score vector $\boldsymbol{s}$, is then set to zero as follows:

$$\boldsymbol{s} = \frac{\partial L_o}{\partial \boldsymbol{\theta}_o} = 0 \,. \tag{21}$$

The elements of the score vector are obtained by taking partial derivatives of the log-likelihood function with respect to each of the parameters:

$$s_\rho = \frac{\partial L_o}{\partial \rho_o} = -\mathrm{tr}(\mathbf{A}_o^{-1}\mathbf{W}_1) + \frac{1}{\sigma_o^2}(\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{B}_o^{\mathrm{T}} \mathbf{G}_o^{-1} \mathbf{B}_o \mathbf{W}_1 \boldsymbol{Y}, \tag{22}$$

$$s_\beta = \frac{\partial L_o}{\partial \beta_o} = \frac{1}{\sigma_o^2}(\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{B}_o^{\mathrm{T}} \mathbf{G}_o^{-1} \mathbf{B}_o \mathbf{X}, \tag{23}$$

and

$$s_{\lambda} = \frac{\partial L_o}{\partial \lambda_o} = -\mathrm{tr}(\mathbf{B}_o^{-1}\mathbf{W}_2) + \frac{1}{\sigma_o^2}(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_2(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)\,, \tag{24}$$

$$s_{\sigma^2} = \frac{\partial L_o}{\partial \sigma_o^2} = -\frac{n}{2\sigma_o^2} + \frac{1}{2\sigma_o^4}(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)\,, \tag{25}$$

$$s_{\gamma} = \frac{\partial L_o}{\partial \gamma_o} = -\frac{1}{2}\mathrm{tr}(\mathbf{D}_o) + \frac{1}{2\sigma_o^2}(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{D}_o\mathbf{B}_o(\mathbf{A}_o Y - \mathbf{X}\boldsymbol{\beta}_o)\,, \tag{26}$$

where $\mathbf{D}_o$ is a diagonal matrix ($n \times n$) with elements:

$$\mathbf{D}_o = \begin{bmatrix} d_{o1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_{on}^2 \end{bmatrix}. \tag{27}$$

The above expressions can be used to derive the estimators of the general model, or of particular cases thereof. Three models are considered in what follows: the linear GWR, GWR with a spatially lagged objective variable (GWR-SL), and GWR with spatial error autocorrelation (GWR-SEA).

**3.2 Linear geographically weighted regression**

The simplest GWR model is obtained by letting $\rho_o = 0$ and $\lambda_o = 0$. In this case $\mathbf{A}_o = \mathbf{I}$, and $\mathbf{B}_o = \mathbf{I}$, which results in the linear model with no spatial association effects:

$$Y = \mathbf{X}\boldsymbol{\beta}_o + \varepsilon_o\,. \tag{28}$$

The covariance matrix of error terms $\boldsymbol{\varepsilon}_o$ is given by equation (18). Estimation and inference of this model have been discussed at length by Brunsdon et al (1996) for the case of global kernel bandwidths, and by Páez et al (2002) for the case of local kernel bandwidths. In addition, Páez et al (2002) derived a test for locational heterogeneity, which implies spatial nonstationarity in the sense defined by Brunsdon et al (1996). In this paper, a complete array of model specification tests are derived in section 5, which include a search for an omitted spatial lag (SL) and spatial error autocorrelation (SEA).

**3.3 GWR with spatially lagged objective variable**

The second model considered results from letting $\lambda_o = 0$, or in other words assuming that there is no spatial error autocorrelation, in which case $\mathbf{B}_o = \mathbf{I}$. In this case, the model becomes the GWR model with a spatially lagged objective variable:

$$Y = \rho_o \mathbf{W}_1 Y + \mathbf{X}\boldsymbol{\beta}_o + \varepsilon_o\,, \tag{29}$$

with covariance matrix again given by equation (18). The above model with local parameters (except for the kernel bandwidth) was proposed by Brunsdon et al (1998) and applied by Páez et al (1999). Previous discussions of this model, however, have applied global kernel bandwidths ($\gamma$ after dropping locational subindex $o$), and dealt with estimation issues while ignoring most inference and hypothesis testing issues. This oversight can be attributed to two important technical limitations: the lack of a method to estimate local kernel bandwidths, and the absence of analytical results to conduct inference of the model. The definition of a general model in this paper provides a way to circumvent the use of global kernel bandwidths, and also the analytical foundation needed to conduct inference and to test the specification of the model (sections 4 and 5).

First, regarding estimation, expressions for the estimators are derived from first-order conditions (23) and (25) as follows:

$$\hat{\boldsymbol{\beta}}_o = (\mathbf{X}^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{A}_o\boldsymbol{Y}, \tag{30}$$

$$\hat{\sigma}_o^2 = \frac{1}{n}(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o). \tag{31}$$

Introducing these expressions into function (19) results in a concentrated log-likelihood that is a function of two parameters, namely $\rho_o$ and $\gamma_o$:

$$L_o^{\mathrm{C}} = -\frac{n}{2}\ln\left[\frac{1}{n}(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_o)\right] - \frac{1}{2}\sum_{i=1}^{n}\gamma_o d_{oi}^2 + \sum_{i=1}^{n}\ln(1-\rho_o w_i), \tag{32}$$

with $w_i$ as the $i$th eigenvalue of matrix $\mathbf{W}_1$ (see Ord, 1975).

As discussed in Páez et al (2002) the method of estimation depends on whether the value of parameter $\gamma_o$ is known. Brunsdon et al (1998) discuss a cross-validation procedure to obtain a global kernel bandwidth that can subsequently be applied as a known constant. An important limitation of this procedure is some preliminary evidence suggesting that cross-validation can result in global estimates of $\gamma$ that may bear little relationship with the underlying spatial variation of the parameter when it exists (Páez et al, 2002). Moreover, because the kernel bandwidth is applied under the assumption that it is a 'known' constant, it cannot be used to test the hypothesis of locational heterogeneity. The natural alternative is to estimate $\gamma_o$ as a parameter of the concentrated log-likelihood (32). In this case, the model has local parameter vector $\boldsymbol{\theta}_o = [\rho_o, \boldsymbol{\beta}_o^{\mathrm{T}}, \sigma_o^2, \gamma_o]^{\mathrm{T}}$ and a total of $3+K$ parameters.

Because the log-likelihood of the model is a function of two parameters, the maximization problem is not as straightforward as for linear GWR models or for the case of global kernel bandwidths. A complex nonlinear optimization procedure could be followed to maximize the log-likelihood simultaneously in $\rho_o$ and $\gamma_o$. In the course of our experiments using the MATLAB computing environment, however, we have found that this procedure sometimes leads to imaginary values of the parameters. A more stable procedure is to iteratively maximize the log-likelihood as a function of one parameter at a time. This is done as follows:

1. *Step 1.* Set $\gamma_o = 0$ as an initial estimate of the parameter and call this value $g_o^s$. Then, obtain an initial estimate of $\rho_o$ by maximizing log-likelihood function (32) as a function of one parameter (that is, the initial estimate $\rho_o$ is the parameter for the global model). Call the solution of this maximization problem $r_o^s$.
2. *Step 2.* Introduce $r_o^s$ into log-likelihood function (32) as an estimate of $\rho_o$ and maximize it as a function of parameter $\gamma_o$. Call the solution of this maximization problem $g_o^{s+1}$.
3. *Step 3.* Introduce $g_o^{s+1}$ into log-likelihood function (32) as an estimate of $\gamma_o$ and maximize it as a function of parameter $\rho_o$. Call the solution of this maximization problem $r_o^{s+1}$.
4. *Step 4.* If $\mathrm{abs}(r_o^s - r_o^{s+1}) > \tau$, or $\mathrm{abs}(g_o^s - g_o^{s+1}) > \tau$, then set $r^s = r^{s+1}$ and return to step 2. Else, proceed to step 5. ($\tau$ is a suitably small tolerance value for convergence, usually $10^{-6}$).
5. *Step 5.* Set $\rho_o = r_o^{s+1}$, and $\gamma_o = g_o^{s+1}$. Using these values, calculate parameter estimates from equations (30) and (31).

We have found that convergence is attained after a reduced number of iterations, usually less than 10 and often 6 or less. Based on the implicit spatial continuity property of the process, the performance of the algorithm could be improved if, for estimation of model $m$, the results of a neighboring local model were used as initial values in step 1 ($g_o^s$ and $r_o^s$) instead of the proposed 0s. This would require of course that the local models be sequentially estimated in a spatially logical order. MATLAB code to estimate the above local model is available from the first author upon request.

### 3.4 GWR with spatial error autocorrelation

The third model considered ensues from letting $\rho_o = 0$, in which case $\mathbf{A}_o = \mathbf{I}$. The result is the GWR with spatial error autocorrelation:

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}_o + \boldsymbol{\varepsilon}_o \,, \tag{34}$$

$$\boldsymbol{\varepsilon}_o = \lambda_o \mathbf{W}_2 \boldsymbol{\varepsilon}_o + \boldsymbol{\mu}_o \,, \tag{35}$$

with covariance matrix (18). As before, the first step towards estimation is to derive expressions for the estimators from first-order conditions (23) and (25) as follows:

$$\hat{\boldsymbol{\beta}}_o = (\mathbf{X}^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\boldsymbol{Y}, \tag{36}$$

$$\hat{\sigma}_o^2 = \frac{1}{n}(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o(\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}_o)\,. \tag{37}$$

Introducing these expressions into function (19) results in a concentrated log-likelihood that is a function of two parameters, namely $\lambda_o$ and $\gamma_o$:

$$L_o^{\mathrm{C}} = -\frac{n}{2}\ln\left[\frac{1}{n}(\boldsymbol{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_o)^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o(\boldsymbol{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_o)\right] - \frac{1}{2}\sum_{i=1}^{n}\gamma_o d_{oi}^2 + \sum_{i=1}^{n}\ln(1-\lambda_o w_i)\,, \tag{38}$$

with $w_i$ as the $i$th eigenvalue of matrix $\mathbf{W}_2$. Estimation of local model parameters, including $\gamma_o$, could follow an iterative procedure similar to that discussed in section 3.3.

## 4 Inference

### 4.1 Asymptotic covariance matrix for the full model

To calculate the asymptotic covariance matrix, needed for hypothesis testing, the elements of the information matrix must first be obtained. These are the negative expected values of the second partial derivatives of the various parameter combinations, after applying the following expressions that arise from the structure of the error terms:

$$\mathrm{E}(\boldsymbol{Y}) = \mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o \,, \tag{39}$$

and

$$\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}) = (\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}(\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o) + \frac{1}{\sigma_o^2}\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}\mathbf{G}_o^{-1}\mathbf{A}_o^{\mathrm{T}-1}\mathbf{B}_o^{\mathrm{T}-1}\,, \tag{40}$$

The element $F_{\rho\rho}$ of information matrix $\mathbf{F}_{\theta\theta}$, for instance, is obtained as follows from the second derivative of the likelihood function with respect to $\rho_o$, and application of the trace operator to give a form that can be transformed into an expectation:

$$\frac{\partial^2 L_o}{\partial \rho_o^2} = -\mathrm{tr}(\mathbf{A}_o^{-1}\mathbf{W}_1)^2 - \frac{1}{\sigma_o^2}\mathrm{tr}(\mathbf{B}_o\mathbf{W}_1)^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\mathbf{W}_1\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}\,. \tag{41}$$

Because the individual elements of the information matrix are the negative expected values of the second derivative, in this case equation (41), it follows that (the first term being nonstochastic):

$$F_{\rho\rho} = -\mathrm{E}\left(\frac{\partial^2 L_o}{\partial \rho^2}\right) = \mathrm{tr}(\mathbf{A}_o^{-1}\mathbf{W}_1)^2 + \frac{1}{\sigma_o^2}\mathrm{tr}(\mathbf{B}_o\mathbf{W}_1)^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\mathbf{W}_1\,\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}})$$

$$= \mathrm{tr}(\mathbf{A}_o^{-1}\mathbf{W}_1)^2 + \mathrm{tr}\mathbf{G}_o(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1})^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1})$$

$$+ \frac{1}{\sigma_o^2}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o)\,. \tag{42}$$

Derivation of the elements of the information matrix is a straightforward but rather lengthy procedure. The rest of the elements appear in appendix A (also see Anselin, 1988a).

Once the elements of matrix $\mathbf{F}_{\theta\theta}$ have been determined, the asymptotic covariance matrix can be obtained as follows:

$$\mathbf{V}_{\theta\theta} = \mathbf{F}_{\theta\theta}^{-1}\,. \tag{43}$$

The variance, standard errors, and $t$-values needed to test the significance of individual parameters are obtained as usual from the diagonal elements of the above matrix. In addition, the information matrix is used to test hypotheses regarding the specification of the model as discussed in section 5.

Two further issues that arise in local inference based on GWR deserve comment. The first is that of simultaneous inference. Briefly stated, the problem is that, whenever hypotheses for more than one local model ($m > 1$) are tested (for example, parameter significance, model specification), protection should be provided to the group of statements through group error rates (Miller, 1981). Anselin (1995) and Ord and Getis (1995; 2001) proposed the use of a simple Bonferroni adjustment for simultaneous testing in the context of local statistics of spatial autocorrelation. This solution, although technically correct, tends to be overconservative for highly correlated (that is, dependent) tests. To overcome this difficulty, Páez et al (2002) introduced a multistep rejective procedure based on the Bonferroni inequality that explicitly takes into account the lack of independence among tests. We refer the reader to that paper for a more in depth discussion of the topic.

Finally, it should be noted that the results derived in this section are asymptotic and therefore valid for large samples. Regarding the application of asymptotic results to small spatial samples, Haining noted in 1990 (page 137) that there are few reported studies on distributional behavior, and the situation does not seem to have changed since. With the exception of two papers exploring the small sample properties of tests for spatial association in regression models (Anselin and Florax, 1995; Anselin and Rey, 1991) to the best of our knowledge not much has been done to explore small-sample distributional properties of models with spatial error autocorrelation or a spatially lagged objective variable, let alone heterogeneity. Although this is a matter deserving further study, at this point we can only offer a warning regarding the application of asymptotic results in small samples, typically $n < 80$. Given data availability conditions nowadays, a considerable number of applications are likely to be larger than this by some factor.

**4.2 GWR with spatially objective variable**

The asymptotic covariance matrix for the case of GWR models with spatially lagged objective variable is obtained from the following expression, setting $\lambda_o = 0$ ($\mathbf{B}_o = \mathbf{I}$), and $\rho_o$, $\boldsymbol{\beta}_o$, $\sigma_o^2$, and $\gamma_o$ set to their maximum likelihood estimates, obtained from the method discussed in section 3.3:

$$\mathbf{V}_{\mathrm{GWR-SL}} = \begin{bmatrix} F_{\rho\rho} & & & \\ \boldsymbol{F}_{\beta\rho} & \mathbf{F}_{\beta\beta} & & \\ F_{\sigma^2\rho} & \boldsymbol{0} & F_{\sigma^2\sigma^2} & \\ F_{\gamma\rho} & \boldsymbol{0} & F_{\gamma\sigma^2} & F_{\gamma\gamma} \end{bmatrix}^{-1} . \tag{44}$$

### 4.3 GWR with spatial error autocorrelation

The asymptotic covariance matrix for the case of a GWR model with spatially dependent errors is obtained from the following expression, setting $\rho_o = 0$ ($\mathbf{A}_o = \mathbf{I}$), and $\boldsymbol{\beta}_o$, $\lambda_o$, $\sigma_o^2$, and $\gamma_o$ set to their maximum likelihood estimates:

$$\mathbf{V}_{\mathrm{GWR-SEA}} = \begin{bmatrix} \mathbf{F}_{\beta\beta} & & & \\ \boldsymbol{0} & F_{\lambda\lambda} & & \\ \boldsymbol{0} & F_{\sigma^2\lambda} & F_{\sigma^2\sigma^2} & \\ \boldsymbol{0} & F_{\gamma\lambda} & F_{\gamma\sigma^2} & F_{\gamma\gamma} \end{bmatrix}^{-1} . \tag{45}$$

## 5 Tests

### 5.1 Approaches to testing

In a recent paper Brunsdon et al (1999) explored the issue of spatial error autocorrelation in GWR models. In order to detect the presence of spatial association among the error terms, a situation that can negatively affect the performance of the model and of some important significance testing procedures, they proposed an exploratory approach based on the variogram plot used in geostatistics (Chilès and Delfinier, 1999). The variogram plots the level of covariance between measurements (in this case the errors) at different spatial separations, and when no trend is discernible from the plot, this is taken as an indication of the absence of spatial error autocorrelation. This approach, although useful in a preliminary, exploratory fashion, is limited in two ways. First, the variogram does not constitute a formal diagnostic susceptible to being tested, and depends on the appreciation of the analyst who must evaluate on the basis of a visual inspection of a plot whether spatial autocorrelation is present among the residuals or not. And, second, the approach is of limited practical use because of identifiability issues, as discussed by Brunsdon et al (1999). The problem is that, as the empirical evidence available shows, spatial error autocorrelation can result from a variety of model misspecification situations (see, for example, section 6 below). This means that, even if a clear trend in the variogram plot suggested spatial error autocorrelation, it would not be clear whether this was a result of spatial association or a consequence of heterogeneity effects.

Leung et al (2000) on the other hand, proposed a more formal way of testing for the presence of spatial error autocorrelation in GWR models based on two well-known statistics of spatial association, namely Moran's $I$ and Geary's $c$. Although this approach opens the door to a classical statistical testing situation, it can still be argued that it presents one technical problem and one practical limitation. First, in order to derive the tests it is assumed that the variance is constant under the hypothesis of no spatial autocorrelation. However, Páez et al (2002) argue that this assumption is not adequate because it restricts the applicability of the geographical weighting scheme. A limitation of a more practical sort, on the other hand, is, as above the problem of indentifiability. Because the statistics of spatial association adopted are not model based, one cannot be sure whether the effect, if and when detected, corresponds to genuine spatial error autocorrelation, to an omitted spatial lag, or a combination of these two processes. To complicate the picture further, Anselin has shown with a simple example (1988a, pages 127–129), that Moran's $I$ statistic of spatial association may have some power against heterogeneity, the very effect that GWR tries to model.

Even if the statistic turned out to be significant, we would still be unclear about the source of the problem and possible remedial measures.

A formal, model-based method of testing the significance of a model parameter, built upon the maximum likelihood approach discussed in the preceding sections, makes use of the first-order conditions and the information matrix of the model. The method, known as the Lagrange multiplier test (Breusch and Pagan, 1980; Anselin, 1988b), is attractive because it aims at detecting specific forms of model misspecification and thus overcomes the identifiability limitations associated with the approaches discussed above. Moreover, unlike other likelihood-based approaches (for example, the likelihood ratio, Wald tests), a Lagrange multiplier test can be evaluated by using the results of the restricted model, and thus does not require estimation of the full model.

To define the test, first consider a partitioning of the parameter vector to isolate the parameter(s) of interest as $\boldsymbol{\theta}_o = [\boldsymbol{\theta}_o^{0\mathrm{T}} | \boldsymbol{\theta}_o^{1\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\theta}_o^0$ is a vector of size $\eta \times 1$. The hypothesis that we wish to test is:

$$\mathrm{H}_0\colon \boldsymbol{\theta}_o^0 = \boldsymbol{0}\,. \tag{46}$$

The Lagrange multiplier test, based on a constrained maximization of the log-likelihood function, evaluates the cost of incrementing the number of parameters in the model. The approach starts with a restricted model using parameter vector $\boldsymbol{\theta}_o^1$, and evaluates the cost of moving towards a full model that includes parameter vector $\boldsymbol{\theta}_o^0$ as well. If the test does not reject hypothesis (46) within a given critical range $\alpha$ (usually 0.05), this indicates that the cost of introducing $\boldsymbol{\theta}_o^0$ is too high to be consistent with the data. The test, $\chi^2$ distributed with $\eta$ degrees of freedom, is (see Engle, 1984):

$$\xi = \boldsymbol{s}_{\theta^0}^{\mathrm{T}} \mathbf{V}_{\theta^0\theta^0} \boldsymbol{s}_{\theta^0}\,, \tag{47}$$

with $s_{\theta^0}$ as the corresponding element of the partitioned score vector, and $\mathbf{V}_{\theta^0\theta^0}$ as the corresponding element of the appropriately partitioned asymptotic covariance matrix, both evaluated under the null hypothesis of $\boldsymbol{\theta}_o^0 = \boldsymbol{0}$. In the following sections, we derive an array of tests against several forms of model misspecification.

### 5.2 Specification tests for GWR models

First we deal with specification tests for GWR models. These can be used to confirm the validity of some important modeling assumptions (for example, spatial independence) in the presence of locational heterogeneity (that is, nonstationarity), and to determine whether a given model formulation is a good representation of the data. In what follows, we derive tests for an omitted spatial lag (SL) and spatial error autocorrelation (SEA) in linear GWR models, and a test for spatial error autocorrelation in GWR models with spatially lagged objective variable.

#### 5.2.1 *Omitted spatial lag in GWR models*

The restricted model in this case is the linear GWR specification [equation 28)], and we wish to test the hypothesis of spatial independence of the objective variable, while assuming error autocorrelation away ($\lambda_o = 0$):

$$\mathrm{H}_0\colon \theta_o^0 = \rho_o = 0\,. \tag{48}$$

In this case, the partitioned vector is $\boldsymbol{\theta}_o = [\rho_o | \boldsymbol{\beta}_o^{\mathrm{T}}, \sigma_o^2, \gamma_o]^{\mathrm{T}} = [\theta_o^0 | \boldsymbol{\theta}_o^{1\mathrm{T}}]^{\mathrm{T}}$, and $\eta = 1$. To derive the test statistic, the appropriately partitioned score vector is obtained from the first-order condition corresponding to parameter $\rho_o$ evaluated under the null hypothesis ($\rho_o = 0$) and parameters $\boldsymbol{\beta}_o$, $\sigma_o^2$, and $\gamma_o$ set to their maximum likelihood estimates:

$$s_\rho = \frac{1}{\sigma_o^2} (\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{G}_o^{-1} \mathbf{W}_1 \boldsymbol{Y}\,. \tag{49}$$

The covariance matrix for the full model is (44), evaluated under the null and with other parameters set to the maximum likelihood estimators of the restricted model:

$$\mathbf{V}_{\theta\theta} = \left[\begin{array}{c|ccc} \boldsymbol{F}_{\rho\rho} & & & \\ \hline \mathbf{F}_{\beta\rho} & \mathbf{F}_{\beta\beta} & & \\ 0 & \boldsymbol{0} & F_{\sigma^2\sigma^2} & \\ 0 & \boldsymbol{0} & F_{\gamma\sigma^2} & F_{\gamma\gamma} \end{array}\right]^{-1} = \begin{bmatrix} V_{\theta^0\theta^0} & \mathbf{V}_{\theta^0\theta^1} \\ \boldsymbol{V}_{\theta^1\theta^0} & \mathbf{V}_{\theta^1\theta^1} \end{bmatrix}. \qquad (50)$$

Given the block diagonal structure of the covariance matrix for the full model evaluated under the null, the expression for the partitioned inverse simplifies considerably to:

$$V_{\theta^0\theta^0} = (F_{\rho\rho} - \boldsymbol{F}_{\rho\beta}\mathbf{F}_{\beta\beta}^{-1}\boldsymbol{F}_{\rho\beta}^{\mathrm{T}})^{-1}. \qquad (51)$$

Substituting equations (49) and (51) in equation (47) gives the Lagrange multiplier test for an omitted spatial lag in GWR models:

$$\xi_{\mathrm{SL}} = \left[\frac{1}{\sigma_o^2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_1\boldsymbol{Y}\right]^2 (F_{\rho\rho} - \boldsymbol{F}_{\rho\beta}\mathbf{F}_{\beta\beta}^{-1}\boldsymbol{F}_{\rho\beta}^{\mathrm{T}})^{-1}. \qquad (52)$$

The above statistic is $\chi^2$ distributed with 1 degree of freedom.

5.2.2 *Spatial error autocorrelation in GWR models*

The second test to be considered regards the presence of spatial error autocorrelation (SEA) in linear GWR models. In this case the full model is (34)–(35), with parameter vector $\boldsymbol{\theta}_o = [\lambda_o\,|\,\boldsymbol{\beta}_o^{\mathrm{T}}, \sigma_o^2, \gamma_o]^{\mathrm{T}} = [\theta_o^0|\boldsymbol{\theta}_o^{1\mathrm{T}}]^{\mathrm{T}}$, and we wish to test the hypothesis of no spatial error autocorrelation while assuming away other spatially association effects (that is, $\rho_o = 0$). Now, the incumbent hypothesis is:

$$\mathrm{H}_0\colon \boldsymbol{\theta}_o^0 = \lambda_o = 0. \qquad (53)$$

The appropriately partitioned score vector is given by the first-order condition corresponding to parameter $\lambda_o$ evaluated under the null hypothesis and parameters $\boldsymbol{\beta}_o$, $\sigma_o^2$, and $\gamma_o$ set to their maximum likelihood estimates:

$$s_\lambda = \frac{1}{\sigma_o^2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_2(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o). \qquad (54)$$

The covariance matrix for the full model is (45), as before evaluated under the null and with parameters of the restricted model $\boldsymbol{\theta}_o^1$ set to their maximum likelihood estimates:

$$\mathbf{V}_{\theta\theta} = \left[\begin{array}{c|ccc} F_{\lambda\lambda} & & & \\ \hline \boldsymbol{0} & \mathbf{F}_{\beta\beta} & & \\ 0 & \boldsymbol{0} & F_{\sigma^2\sigma^2} & \\ 0 & \boldsymbol{0} & F_{\gamma\sigma^2} & F_{\gamma\gamma} \end{array}\right]^{-1} = \begin{bmatrix} V_{\theta^0\theta^0} & \boldsymbol{V}_{\theta^0\theta^1} \\ \boldsymbol{V}_{\theta^1\theta^0} & \mathbf{V}_{\theta^1\theta^1} \end{bmatrix}. \qquad (55)$$

Again, the structure of the covariance matrix is block diagonal, which considerably simplifies matters, and the statistic ($\chi^2$ distributed with $\eta = 1$ degree of freedom) can be expressed as:

$$\xi_{\mathrm{SEA}} = \left[\frac{1}{\sigma_o^2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_2(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)\right]^2 (\mathrm{tr}\mathbf{W}_2^2 + \mathrm{tr}\mathbf{G}_o\mathbf{W}_2^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_2)^{-1}. \qquad (56)$$

It is easy to verify that statistic (56) is a special form of the statistic for spatial error autocorrelation in the presence of heteroscedasticity derived by Anselin (1988a, pages 106–107).

5.2.3 *Spatial error autocorrelation in GWR with spatially lagged objective variable*
The restricted model in this case is the GWR specification with a spatially lagged objective variable, and the parameter vector becomes

$$\boldsymbol{\theta}_o = [\lambda_o | \rho_o, \boldsymbol{\beta}_o^{\mathrm{T}}, \sigma_o^2 \gamma_o]^{\mathrm{T}} = [\theta_o^0 | \boldsymbol{\theta}_o^{1\mathrm{T}}]^{\mathrm{T}}.$$

The hypothesis of interest is ($\eta = 1$):

$$\mathrm{H}_0\colon \theta_o^0 = \lambda_o = 0. \tag{57}$$

Following the procedures outlined above, we can show the corresponding statistic to be:

$$\xi_{\mathrm{SEA2}} = \left[\frac{1}{\sigma_o^2}(\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{G}_o^{-1} \mathbf{W}_2 (\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)\right]^2 (F_{\lambda\lambda} - F_{\lambda\rho}^2 V_{\rho\rho})^{-1}. \tag{58}$$

In this case, the resulting expression is more complex than the previous statistics. However, it can be readily calculated with elements available from the estimation of the restricted model. $V_{\rho\rho}$, for example, is the corresponding element of the asymptotic covariance matrix (44). The above statistics can be used to verify that a GWR model is well specified and complies with the assumptions laid upon it.

### 5.3 Specification tests for global models

Other situations of practical interest demand that we deal with specification tests for global models. In applied research, the analyst may wish to work on a relatively simpler global model, and to test it for the presence of spatial effects before proceeding to estimate more complex local models. The tests for an omitted spatial lag (SL), and for spatial error autocorrelation (SEA) in global models derived by Anselin (1988b) have been widely applied in spatial econometrics model specification searches. We will next derive two statistics to test for the presence of locational heterogeneity (LH, or in other words, spatial nonstationarity) in global models that include spatial association components.

5.3.1 *Locational heterogeneity in models with a spatially lagged objective variable*
The restricted model in this case is the global model with spatially lagged objective variable [equation (29) assuming that $\gamma_o = 0$ and $\lambda_o = 0$], and the hypothesis we wish to test is:

$$\mathrm{H}_0\colon \boldsymbol{\theta}_o^0 = \gamma_o = 0. \tag{59}$$

Clearly, the test for locational heterogeneity is a specification test for a spatial form of nonconstant variance. In this case $\eta = 1$ and the appropriately partitioned score vector is given by the first-order condition corresponding to parameter $\gamma_o$ evaluated under the null ($\gamma_o = 0$) and parameters $\rho_o$, $\boldsymbol{\beta}_o$, and $\sigma_o^2$ set to their maximum likelihood estimates:

$$s_\gamma = \frac{1}{2}\left[\frac{1}{\sigma_o^2}(\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}} \mathbf{D}_o (\mathbf{A}_o \boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o) - \mathrm{tr}\mathbf{D}_o\right]. \tag{60}$$

The covariance matrix for the full model is (44), after rearranging rows and columns to allow for a partitioning of the matrix that isolates $\theta_o^0$:

$$\mathbf{V}_{\theta\theta} = \left[\begin{array}{c|ccc} F_{\gamma\gamma} & & & \\ \hline F_{\rho\gamma} & F_{\rho\rho} & & \\ \boldsymbol{0} & \boldsymbol{F}_{\beta\rho} & \mathbf{F}_{\beta\beta} & \\ F_{\sigma^2\gamma} & F_{\sigma^2\rho} & \boldsymbol{0} & F_{\sigma^2\sigma^2} \end{array}\right]^{-1} = \begin{bmatrix} V_{\theta^0\theta^0} & \boldsymbol{V}_{\theta^0\theta^1} \\ \boldsymbol{V}_{\theta^1\theta^0} & \mathbf{V}_{\theta^1\theta^1} \end{bmatrix}, \tag{61}$$

so that $V_{\theta^0\theta^0}$ becomes:

$$V_{\theta^0\theta^0} = (F_{\gamma\gamma} - [F_{\gamma\rho} \quad \boldsymbol{0} \quad F_{\gamma\sigma^2}]\mathbf{V}_{\theta^1\theta^1}[F_{\gamma\rho} \quad \boldsymbol{0}^{\mathrm{T}} \quad F_{\gamma\sigma^2}]^{\mathrm{T}})^{-1}. \tag{62}$$

$\mathbf{V}_{\theta^1\theta^1}$ is simply the asymptotic covariance matrix of the restricted model (the global SL model), and the rest of the elements are evaluated under the null hypothesis of $\gamma_o = 0$. Substituting equations (62) and (60) in equation (47) gives the Lagrange multiplier test for locational heterogeneity in a global model with a spatially lagged objective variable:

$$\xi_{\mathrm{LH}} = \frac{1}{4}\left[\frac{1}{\sigma_o^2}(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{D}_o(\mathbf{A}_o\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o) - \mathrm{tr}\mathbf{D}_o\right]^2 V_{\theta^0\theta^0} \,. \tag{63}$$

The above statistic is $\chi^2$ distributed with 1 degree of freedom.

5.3.2 *Locational heterogeneity in the presence of spatial error autocorrelation*
In this case, the restricted model is the global specification with spatial error autocorrelation [SEA; equations (34) and (35), assuming $\gamma_o = 0$ and $\rho_o = 0$], and the hypothesis we wish to test is again:

$$\mathrm{H}_0\colon \theta_o^0 = \gamma_o = 0 \,.$$

Now the appropriately partitioned score vector is given by the first-order condition corresponding to parameter $\gamma_o$ evaluated under the null ($\gamma_o = 0$) and parameters $\boldsymbol{\beta}_o$, $\lambda_o$, and $\sigma_o^2$ set to their maximum likelihood estimates:

$$s_\gamma = \frac{1}{2}\left[\frac{1}{\sigma_o^2}(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{B}_o{}^{\mathrm{T}}\mathbf{D}_o\mathbf{B}_o(\boldsymbol{Y} - \mathbf{X}\boldsymbol{\beta}_o) - \mathrm{tr}\mathbf{D}_o\right] . \tag{64}$$

The covariance matrix for the full model is (45), after rearranging the terms:

$$\mathbf{V}_{\theta\theta} = \left[\begin{array}{c:ccc} F_{\gamma\gamma} & & & \\ \hdashline \boldsymbol{0} & \mathbf{F}_{\beta\beta} & & \\ F_{\lambda\gamma} & \boldsymbol{0} & \mathbf{F}_{\lambda\lambda} & \\ F_{\sigma^2\gamma} & \boldsymbol{0} & F_{\sigma^2\gamma} & F_{\sigma^2\sigma^2} \end{array}\right]^{-1} = \begin{bmatrix} V_{\theta^0\theta^0} & \boldsymbol{V}_{\theta^0\theta^1} \\ \mathbf{V}_{\theta^1\theta^0} & \boldsymbol{V}_{\theta^1\theta^1} \end{bmatrix}, \tag{65}$$

so that $V_{\theta^0\theta^0}$ becomes:

$$V_{\theta^0\theta^0} = (F_{\gamma\gamma} - [\boldsymbol{0} \;\; F_{\lambda\gamma} \;\; F_{\gamma\sigma^2}]\mathbf{V}_{\theta^1\theta^1}[\boldsymbol{0}^{\mathrm{T}} \;\; F_{\lambda\gamma} \;\; F_{\gamma\sigma^2}]^{\mathrm{T}})^{-1} \,. \tag{66}$$

$V_{\theta^1\theta^1}$ is simply the asymptotic covariance matrix of the restricted model (the global model with SEA), and the rest of the elements are evaluated under the null hypothesis of $\gamma_o = 0$ and other parameters set to their maximum likelihood estimates. Substituting equations (66) and (64) in equation (47) gives the Lagrange multiplier test for locational heterogeneity in global models with spatial error autocorrelation (also see Anselin, 1988a; pages 120–124). Application of these tests is exemplified in the following section.

## 6 Application: urban heat island in Sendai, revisited

### 6.1 Global models

In this section we apply the results obtained above regarding the tests for locational heterogeneity in global models. To this end we revisit the problem of the urban heat island in Sendai City, Japan (Páez et al, 1998; 2002). We begin our example by noting that the purpose of this application is not to offer a comprehensive analysis of the problem, but rather to illustrate a number of points that we gather to be of interest in the application of the GWR method. The variable we wish to explain is the temperature ($T$ in Kelvin) at different locations of the city during a heat island episode by using variables introduced to represent the physical characteristics of the city. These include a physical measure of city size, given by the natural logarithm of distance from the center of the city to the location of the observation or (LDIST), and three land-use variables defined in terms of the percentage of cover by zone. The latter variables represent three main types of land use, namely commercial (C%), industrial (I%), and residential (R%) land uses. These variables are indicators of both land-use intensity by type and the spatial distribution of uses.

First, using the above variables, we estimate global models (that is, models that assume constant variance), to produce the results shown in table 1. The first model estimated is a linear regression with no spatial association components. Estimation and specification tests for these models use a row-standardized matrix of spatial interactions defined on a nearest neighbor criterion, and letting $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}_{st}$. The tests show that the linear model is an inadequate representation of the data: the trio of Lagrange multiplier tests for an omitted spatial lag (SL), spatial error autocorrelation (SEA), and locational heterogeneity [LH, see Páez et al (2002); the value reported is the maximum among 493 local tests] all turn out to be significant. The level of significance is at least 0.05, although it is important to note that, in the case of the locational heterogeneity test, said value is adjusted for simultaneous inference, according to the procedure discussed in Páez et al (2002). It is interesting to note that testing for spatial error autocorrelation by means of Moran's $I$ statistic also yields a significant result. However, besides detecting the effect in general, the test offers little clue as to the source of the problem or any possible remedial measures, thus underscoring the problem of identifiability discussed in section 5.

In the current example, the size of the tests for the linear model seems to suggest that an omitted spatial lag is the dominant effect (see Anselin and Rey, 1991). In order to verify this suspicion, we proceed to estimate global models with a spatially lagged objective variable (SL) and with spatial error autocorrelation SEA. These models are estimated and tested with the methods discussed in Anselin (1988a), and the results are shown in columns 3 and 4 of table 1. Testing the specification of the SL model shows that spatial error autocorrelation is no longer an issue, although the use of statistic (63) shows that there is significant locational heterogeneity, thus suggesting that GWR-SL would be a better model specification. The model with SEA on the other hand, is completely unsatisfactory, because it fails to solve lingering problems with an

**Table 1.** Global models—estimation results.

| Parameter | Linear model estimate | SL model estimate | SEA model estimate |
|---|---|---|---|
| $\rho$ | | 0.85966* | |
| CONST | 308.53675* | 42.52935* | 308.20376* |
| LDIST | −0.88662* | −0.04832 | −0.80663* |
| C% | 0.00206 | 0.00135 | 0.00121 |
| I% | 0.00558 | 0.00322 | 0.00234 |
| R% | 0.02333* | 0.00955* | 0.00970* |
| $\lambda$ | | | 0.87101* |
| $\sigma^2$ | 1.136 | 0.365 | 0.367 |
| $R^2_{adj}$ | 0.418 | | |
| Log-likelihood | −560.123 | −329.626 | −332.87 |
| Spatial association tests | | | |
| Lagrange multipliers | | | |
| SL | 580.774* | | 3886.973* |
| SEA | 586.799* | 0.190 | |
| LH (maximum) | 224.5114* | 165.8724* | 164.6601* |
| Normalized Moran's $I$ | | | |
| $Z(I)$ | 24.262* | | |

Note. SL spatial lag, SEA spatial error autocorrelation, LH locational heterogeneity.
* Significance 0.05.

omitted spatial lag and locational heterogeneity. It was thus decided to continue the analysis estimating and testing GWR models with spatially lagged dependent variable, using the methods discussed in section 3.3.

### 6.2 Spatially autoregressive GWR: Estimation and inference

Next, we estimate and test local GWR-SL models. We conduct estimation at every location corresponding to a recorded observation, to give a total of $m = 493$ local models. The first step to estimate a model is to obtain initial parameter estimates for $\rho_o$ and $\gamma_o$. As discussed above, a way to obtain initial values is to estimate a global SL

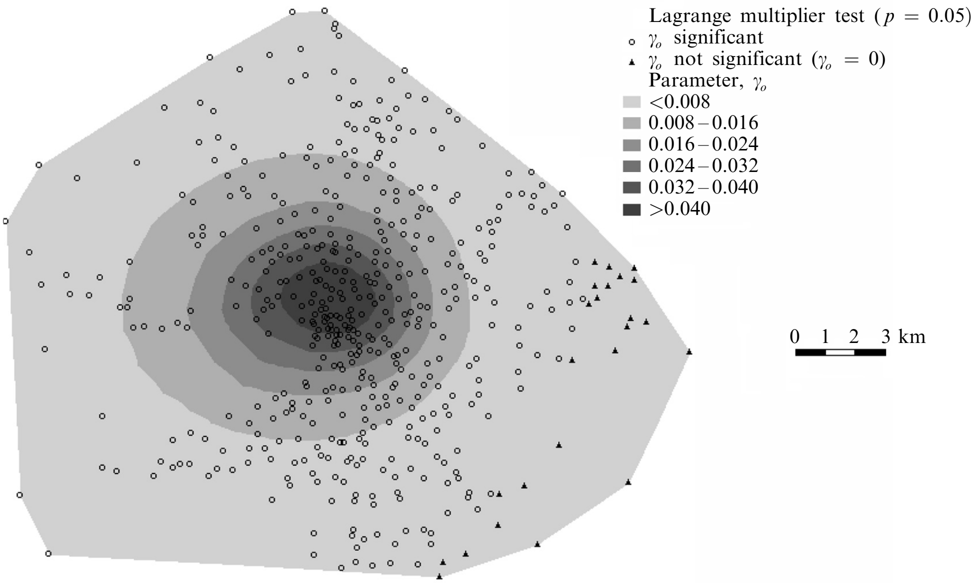


**Figure 1.** Spatial distribution of parameter $\gamma_o$ and significance tests.

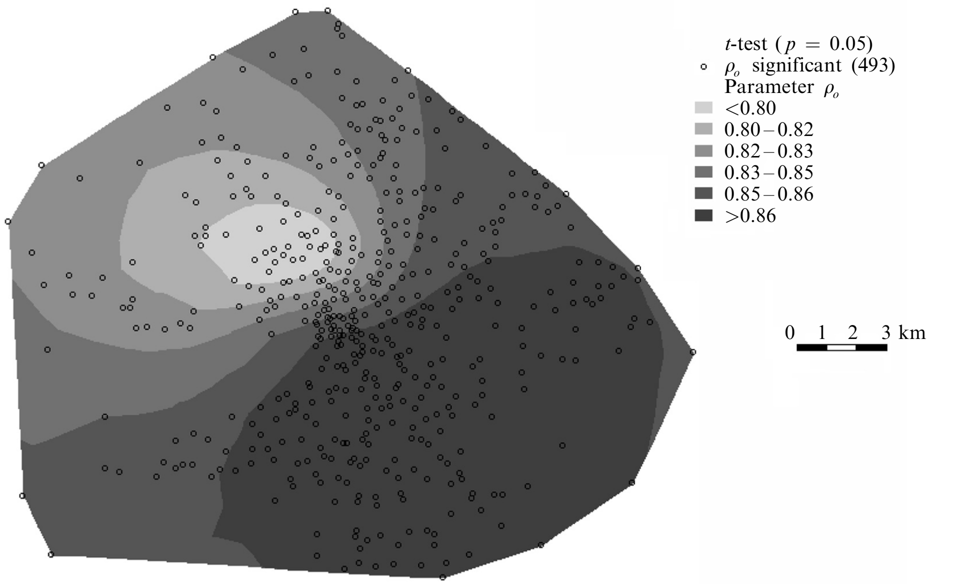


**Figure 2.** Spatial distribution of parameter $\rho_o$ and significance tests.

model, thus effectively setting $\gamma_o = 0$. The value of $\rho_o$ obtained in this fashion (in this case 0.859; see table 1) is then used to maximize log-likelihood function (32) to obtain a new estimate for $\gamma_o$, which is in turn used to maximize the log-likelihood in $\rho_o$. For each local model, this procedure is repeated until convergence is achieved.

The results of estimating these local models appear in figures 1–4. First, it can be seen in figure 1 that parameter $\gamma_o$ ranges between 0.002 (the smallest *significant* kernel bandwidth) and 0.0486, and is significant for most of the study area, with the exception of 24 locations in the southeast region of the city. The significance of the parameter is tested by means of the Lagrange multipliers approach [equation (63)] with a nominal level of significance of 0.05 corrected by means of the multiple inference procedure discussed in Páez et al (2002). From the figure it can be seen that spatial

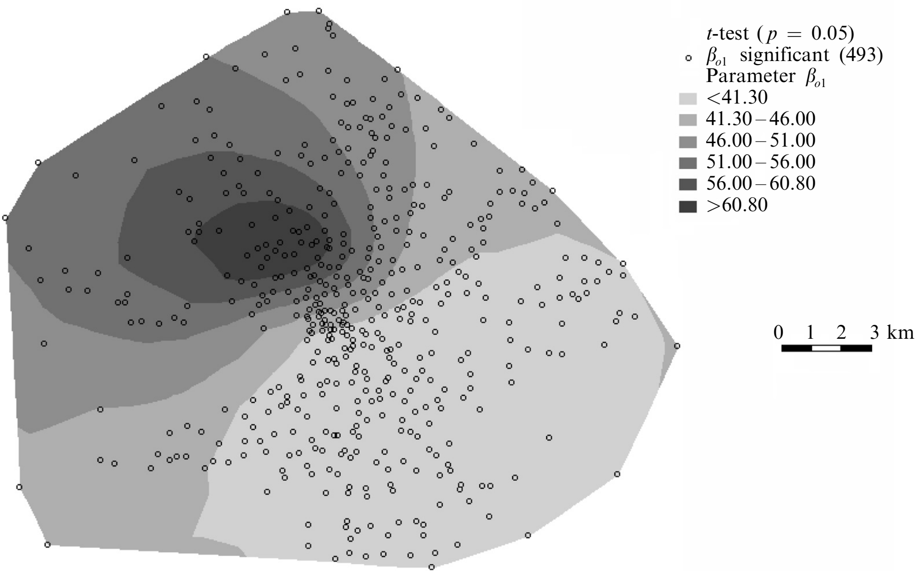


**Figure 3.** Spatial distribution of parameter $\beta_{o1}$ (CONST) and significance tests.

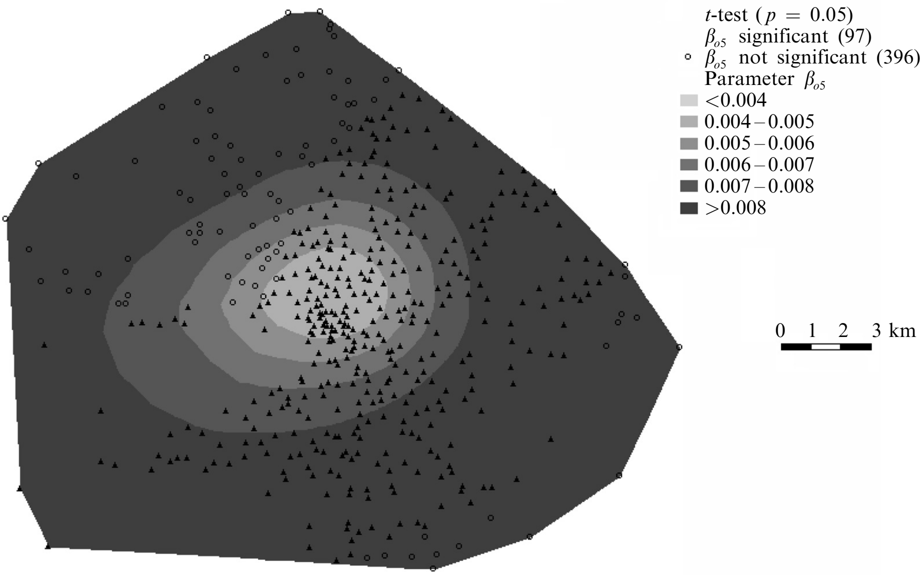


**Figure 4.** Spatial distribution of parameter $\beta_{o5}$ (R%) and significance tests.

variation of parameter $\gamma_o$ is rather regular, thus suggesting that the underlying process is continuous but not uniform across space. Comparing these results with those obtained for linear GWR models (see Páez et al, 2002) it is interesting to note that the degree of locational heterogeneity decreases considerably once a spatially lagged objective variable has been introduced. This suggests that spatial error autocorrelation detected in the global linear model is caused jointly by an omitted spatial lag and heterogeneity.

Another result of introducing these two spatial effects into the model specification is that parameter significance levels also change. For instance, parameter $\rho$ and the parameters associated with variables LDIST, and R% were significant in the global SL model. However, after estimating local GWR-SL models, R% ceases to be significant in a large number of locations (see figure 4). A well-known consequence of ignoring heteroscedasticity in regression models is potentially misleading significance levels, which seems to be the case with the global model in the present example. Estimation results appear in figures 2–4, along with their corresponding significance tests. It can be seen there that there is a considerable degree of spatial variation among local parameters: 12.14% between the maximum and minimum local $\rho_o$, 79.48% for the constant, and 121.54% for residential land uses; this variation was invisible to the global version of the model. A final test of the local models is conducted to detect the presence of spatial error autocorrelation, using statistic (58). The maximum value for this statistic (from among 493 locations) is 3.723, clearly a nonsignificant result at the 0.05 level, even without considering the adjustment for simultaneous hypothesis testing. This is evidence that the local models are correctly specified.

### 6.3 Model comparison

In this section we compare in a rather informal way the performance of five different modeling schemes, namely the global linear model, the linear GWR using local kernel bandwidths and spatially invariant (global) kernel bandwidths (see Páez et al, 2002), and two of the approaches considered in the preceding sections: the global model with a spatially lagged objective variable, and GWR-SL. Brunsdon et al (1999) have proposed a more formal method of assessing the goodness of fit of GWR models based on the analysis of the variance. Here, we consider two simple indicators. The first one is the sum of squared errors (SSE) obtained from comparing the observed values of the objective variable (temperature $T$) against the values predicted by the corresponding global model or set of local models. And, second, we calculate a pseudo-$R^2$ measure given by the square of the correlation coefficient between the observed values and the values predicted by the models (or the coefficient of determination for the case of the linear global model). The results appear in table 2.

It is clear from the table that in general GWR models produce better fits than their global model counterparts. This result is reasonable when we think that GWR models represent local variability of the process that is overlooked by a global model. In addition to this, there are some interesting insights to be had from the table. First, it turns out that the set of linear GWR models estimated by using a spatially

**Table 2.** Performance indicators.

| Model | SSE | Pseudo-$R^2$ |
|---|---|---|
| 1. Linear model (global) | 560.0821 | 0.418 |
| 2. GWR with local $\gamma_o$ | 484.6498 | 0.527 |
| 3. GWR with constant $\gamma$ | 210.2323 | 0.788 |
| 4. SA model (global) | 179.9987 | 0.819 |
| 5. GWR-SL with local $\gamma_o$ | 176.7668 | 0.823 |

invariant kernel bandwidth gives a better fit than models that estimate local kernel bandwidths. As discussed in Páez et al (2002), in this example the kernel bandwidth obtained from a cross-validation procedure is considerably larger than those obtained from estimating the models locally, including the kernel bandwidth. This in turn means that the estimation 'neighborhood' given by the global kernel bandwidth is smaller, estimation becomes more local, and thus the results tend to resemble more closely the observed response surface with the consequent improvement of fit.

In general, the preliminary evidence available suggests that the use of global kernel bandwidths tends to exaggerate the variability of the process as the bandwidth is selected in an attempt to improve the fit locally. We see merit in this property if the objective of modeling is to obtain better model fits, as, for instance, would be the case when conducting an exercise of spatial interpolation. The interpretation of the technique in such cases would then be closer to the smoothing methods with which GWR has been widely compared. Under such circumstances, the model can be estimated by using well-known methods such as weighted least squares. Unfortunately, our interpretation of the model as a form of spatial error variance heterogeneity forecloses the possibility of testing the hypothesis of locational heterogeneity (and implicitly nonstationarity) when a constant kernel bandwidth is used, although other parameters can be tested in the usual fashion.

Regarding the models with spatially lagged objective variable, it is clear that improvement of the fit obtained from GWR-SL models is marginal at best. The benefits of estimating local models derive in this case from a better understanding of the spatial variability of the process and more reliable inferential statements. These characteristics of the data would have gone unnoticed were it not for the application of statistical tests to diagnose for the whole range of spatial effects, and the modeling options that allowed us to consider them explicit in our example.

As a final indicator of model performance, we plot in figure 5 the local distribution of residuals obtained from three different models, namely the global, the linear GWR, and GWR models with a spatially lagged objective variable. It can be seen there that in

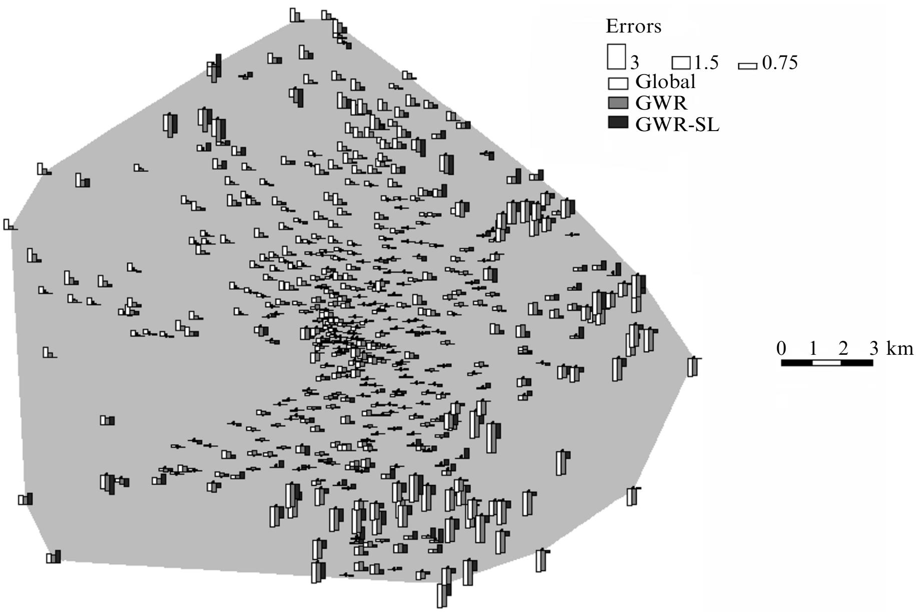


**Figure 5.** Spatial distribution of the errors.

general the models are more accurate towards the center of the study area, but that the quality of the estimates decreases in the direction of the borders, especially towards the southeastern border of the study area. It is not clear at this point to what extent this is a result of so-called border effects, but we suggest that this is a matter deserving further examination.

## 7 Summary and conclusions

In this paper we have presented a generalization of the method to estimate and conduct inference of geographically weighted regression models. This generalization has been made possible by casting the technique within the framework of spatial variance heterogeneity modeling. In this way, we have shown how a general GWR model is in fact a special form of the classical spatial econometrics model discussed by Anselin (1988a). Our results can be summarized as follows:

(1) We demonstrated how the GWR model with spatial association components is a special form of the general spatial econometric model, which results from adopting a variance model that is a continuous function of geographical distance. Also, we showed that the variance function proposed complies with the regularity conditions set out for spatial models, meaning that asymptotic results are valid for the model proposed in this paper.

(2) The general GWR model introduced here covers as particular cases the linear GWR model, as well as GWR with a spatially lagged objective variable, and GWR with spatial error autocorrelation. Based on a likelihood-maximization approach, we discussed the methods to estimate the above models. Moreover, we addressed the inference aspect of these models that was missing from the literature. To this end we derived the analytical results needed to conduct significance tests of the parameters.

(3) We also derived, based on the general model, a complete set of specification tests. The tests are based on the Lagrange multiplier approach and are meant to provide the analyst with a complete set of tools to detect locational heterogeneity in global models, and spatial association effects in GWR models, including spatial error autocorrelation. Availability of these tests means that in practical situations an analyst can be confident about the appropriateness of a model after it has passed all its specification tests, as shown in our case study.

It is our sincere hope that the models and tests presented here will prove a useful addition to the toolbox of applied researchers and scientists dealing with the analysis of spatial data.

**Acknowledgements.** We are indebted to Professor Stewart Fotheringham and Dr Chris Brunsdon for a critical lecture of an earlier version of this paper. We are especially grateful to three unnamed reviewers and to Professor Eric Sheppard for their criticism and comments, which importantly contributed to bringing the paper into focus. We naturally assume full responsibility for any and all errors or shortcomings.

**References**

Anselin L, 1988a *Spatial Econometrics: Methods and Models* (Kluwer, Dordrecht)

Anselin L, 1988b, "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity" *Geographical Analysis* **20** 1 – 17

Anselin L, 1995, "Local indicators of spatial association – LISA" *Geographical Analysis* **27** 93 – 115

Anselin L, Florax R, 1995, "Small sample properties of tests for spatial dependence in regression models", in *New Directions in Spatial Econometrics* Eds L Anselin, R Florax (Springer, Berlin) pp 21 – 74

Anselin L, Rey S, 1991, "Properties of tests for spatial dependence in linear regression models" *Geographical Analysis* **23** 112 – 131

Box G E P, Meyer R D, 1986, "Dispersion effects from fractional design" *Technometrics* **28**(1) 19 – 27

Breusch T S, Pagan A R, 1980, "The Lagrange multiplier test and its applications to model specification in econometrics" *Review of Economic Studies* **48** 239 – 253
Brunsdon C, Fotheringham A S, Charlton M E, 1996, "Geographically weighted regression: a method for exploring spatial nonstationarity" *Geographical Analysis* **28** 281 – 298
Brunsdon C, Fotheringham A S, Charlton M, 1998, "Spatial nonstationarity and autoregressive models" *Environment and Planning A* **30** 957 – 973
Brunsdon C, Fotheringham A S, Charlton M E, 1999, "Some notes on parametric significance tests for geographically weighted regression" *Journal of Regional Science* **39** 497 – 524
Chilès J P, Delfinier P, 1999 *Geostatistics: Modeling Spatial Uncertainty* (John Wiley, New York)
Cliff A D, Ord J K, 1973 *Spatial Autocorrelation* (Pion, London)
Cliff A D, Ord J K, 1981 *Spatial Processes: Models and Applications* (Pion, London)
Engle, R F, 1984, "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics", in *Handbook of Econometrics* Eds Z Griliches, M D Intriligator (Elsevier Science, Amsterdam) pp 775 – 826
Fingleton B, 2000, "Spatial econometrics, economic geography, dynamics and equilibrium: a 'third way'?" *Environment and Planning A* **32** 1481 – 1498
Fotheringham A S, Charlton M, Brundson C, 1997, "Measuring spatial variations in relationships with geographically weighted regression", in *Recent Developments in Spatial Analysis* Eds M M Fischer, A Getis (Springer, Berlin) pp 60 – 82
Fotheringham A S, Charlton M, Brundson C, 1998, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis" *Environment and Planning A* **30** 1905 – 1927
Geary R, 1954, "The contiguity ratio and statistical mapping" *The Incorporated Statistician* **5** 115 – 145
Getis A, Ord J K, 1992, "The analysis of spatial association by use of distance statistics" *Geographical Analysis* **24** 189 – 206
Griffith D, 1988 *Advanced Spatial Statistics* (Kluwer Academic, Dordrecht)
Griffith D, 1999, "Statistical and mathematical sources of regional science theory: map pattern analysis as an example" *Papers in Regional Science* **78** 21 – 45
Griffith D A, Layne L J, 1999 *A Casebook for Spatial Statistical Analysis* (Oxford University Press, Oxford)
Haining R, 1990 *Spatial Data Analysis in the Social and Environmental Sciences* (Cambridge University Press, Cambridge)
Leung Y, Mei C-L, Zhang W-X, 2000, "Testing for spatial autocorrelation among the residuals of the geographically weighted regression" *Environment and Planning A* **32** 871 – 890
Lyon J D, Tsai C L, 1996, "A comparison of tests for heteroscedasticity" *The Statistician* **45** 337 – 349
Miller R G, 1981 *Simultaneous Statistical Inference* 2nd edition (Springer, New York)
Moran P, 1948, "The interpretation of statistical maps" *Journal of the Royal Statistical Society B* **10** 243 – 251
Murdoch J C, Sandler T, 1997, "Voluntary cutbacks and pretreaty behavior: the Helsinki Protocol and sulfur emissions" *Public Finance Review* **25**(2) 139 – 162
Murdoch J C, Sandler T, Sargent K, 1997, "A tale of two collectives: sulfur versus nitrogen oxides emission reduction in Europe" *Economica* **64** 281 – 301
Nass C, Garfinkle D, 1992, "Localized autocorrelation diagnostic statistics (LADS) for spatial models: conceptualization, utilization, and computation" *Regional Science and Urban Economics* **22** 333 – 346
Ord J K, 1975, "Estimation methods for models of spatial interaction" *Journal of the American Statistical Association* **70** 120 – 126
Ord J K, Getis A, 1995, "Local spatial autocorrelation statistics: distributional issues and an application" *Geographic Analysis* **27** 286 – 306
Ord J K, Getis A, 2001, "Testing for local spatial autocorrelation in the presence of global autocorrelation" *Journal of Regional Science* **41** 411 – 432
Páez A, Uchida T, Miyamoto K, 1998, "Urbanization and the urban heat island effect from a spatial descriptive approach", in *Papers on City Planning 33* (The City Planning Institute of Japan, Tokyo) 67 – 72
Páez A, Uchida T, Miyamoto K, 1999, "A spatial analysis of external economies and its association with transportation infrastructure: the case of Sendai City" *Journal of the Eastern Asia Society for Transportation Studies* **3**(4) 149 – 164

Páez A, Uchida T, Miyamoto K, 2001, "Spatial association and heterogeneity issues in land price models" *Urban Studies* **38** 1493 – 1508
Páez A, Uchida T, Miyamoto K, 2002, "A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for spatial nonstationarity" *Environment and Planning A* **34** 733 – 754
Verbyla A P, 1993, "Modeling variance heterogeneity: residual maximum likelihood and diagnostics" *Journal of the Royal Statistical Society B* **55** 493 – 508

## APPENDIX

The elements of the information matrix are the negative expected values of the second partial derivatives of the various parameter combinations, after applying the expected values in equations (39) – (40). These can be shown to be:

$$F_{\rho\rho} = \mathrm{tr}(\mathbf{A}_o^{-1}\mathbf{W}_1)^2 + \mathrm{tr}\mathbf{G}_o(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1})^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}) + \frac{1}{\sigma_o^2}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o)^{\mathrm{T}}\mathbf{G}_o^{-1}(\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o), \tag{A1}$$

$$\boldsymbol{F}_{\beta\rho} = \frac{1}{\sigma_o^2} - (\mathbf{B}_o\mathbf{X})^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{X}\boldsymbol{\beta}_o, \tag{A2}$$

$$\mathbf{F}_{\beta\beta} = \frac{1}{\sigma_o^2} - \mathbf{X}^{\mathrm{T}}\mathbf{B}_o^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{X}\mathbf{B}_o, \tag{A3}$$

$$F_{\lambda\rho} = \mathrm{tr}(\mathbf{W}_2\mathbf{B}_o^{-1})^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}\mathbf{G}_o + \mathrm{tr}(\mathbf{W}_2\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}), \tag{A4}$$

$$F_{\lambda\lambda} = \mathrm{tr}(\mathbf{W}_2\mathbf{B}_o^{-1})^2 + \mathrm{tr}\mathbf{G}_o(\mathbf{W}_2\mathbf{B}_o^{-1})^{\mathrm{T}}\mathbf{G}_o^{-1}\mathbf{W}_2\mathbf{B}_o^{-1}, \tag{A5}$$

$$F_{\sigma^2\rho} = \frac{1}{\sigma_o^2}\mathrm{tr}\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}, \tag{A6}$$

$$F_{\sigma^2\lambda} = \frac{1}{\sigma_o^2}\mathrm{tr}\mathbf{W}_2\mathbf{B}_o^{-1}, \tag{A7}$$

$$F_{\sigma^2\sigma^2} = \frac{n}{2\sigma_o^2}, \tag{A8}$$

$$F_{\gamma\rho} = \mathrm{tr}\mathbf{D}_o\mathbf{B}_o\mathbf{W}_1\mathbf{A}_o^{-1}\mathbf{B}_o^{-1}, \tag{A9}$$

$$\boldsymbol{F}_{\lambda\beta} = \boldsymbol{F}_{\sigma^2\beta} = \boldsymbol{F}_{\gamma\beta} = \boldsymbol{0}, \tag{A10}$$

$$F_{\lambda\gamma} = \mathrm{tr}\mathbf{D}_o\mathbf{W}_2\mathbf{B}_o^{-1}, \tag{A11}$$

$$F_{\gamma\sigma^2} = \frac{1}{2\sigma_o^2}\mathrm{tr}\mathbf{D}_o, \tag{A12}$$

$$F_{\gamma\gamma} = \frac{1}{2}\mathrm{tr}\mathbf{D}_o^2. \tag{A13}$$