Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Transport Geography 14 (2006) 167–176
JOURNAL OF TRANSPORT GEOGRAPHY
www.elsevier.com/locate/jtrangeo

Exploring contextual variations in land use and transport analysis using a probit model with geographical weights

Antonio Páez *

Centre for Spatial Analysis/School of Geography and Earth Sciences, McMaster University, 1280 Main Street West, Hamilton, Ont., Canada L8S 1A1

Abstract

A majority of statistical methods used in the analysis of land use and transportation systems implicitly carry the assumption that relationships are constant across locations or individuals, thus ignoring contextual variation due to geographical or socio-economic heterogeneity. In some cases, where the assumption of constant relationships is questionable, market segmentation procedures are used to study varying relationships. More recently, methodological developments, and a greater awareness of the importance of geography, have led to increasingly sophisticated ways to explore varying relationships in land use and transportation modeling. The objective of this paper is to propose a simple probit model to explore contextual variability in continuous-space. Some conceptual and technical issues are discussed, and an example is presented that reanalyzes land use change using data from California's BART system. The results of the example suggest that considerable parametric variation exists across geographical space, and thus underlines the relevance of contextual effects.

Keywords: Heteroscedastic probit model; Geographically weighted regression; Variance functions; Maximum likelihood estimation; Transportation and land use analysis

1. Introduction

The analysis of land use and transportation systems often relies on the use of statistical or econometric models for explanatory or forecasting purposes. Regression analysis, for example, is a technical element for the study of trip generation, trip distribution and modal split in the standard 4-stages modeling approach (McNally, 2000b). Other aspects of travel behavior, including modal choice and route choice are studied using discrete choice models—that is, limited dependent variable models embedded in an economic utility maximization framework. Models of this type can be used to study travel behavior from the perspective of activities in the activity-based analysis paradigm (McNally, 2000a). Locational analysis, such as applied in land use models, also makes use of discrete choice models (e.g. Martínez, 1992; Miyamoto, 1993), and land use change has been studied using limited dependent variable models (e.g. Landis et al., 1995; McMillen, 1989). Discrete choice models and limited dependent variable models share the same technical basis. They also share some common simplifications, including the assumption of stationary (i.e. stable) relationships between variables. Stationary relationships, in turn, ignore the possibility of local variation (i.e. contextual effects) due to heterogeneity.

The assumption of stationary relationships is useful to obtain relatively simple and easily estimable models. Such models, however, may be of limited value when the empirical evidence does not support the assumption. In other cases, the assumption itself may be inconsistent with given theoretical propositions. In geographical analysis, for example, a candidate principle for the second law of geography is that of spatial heterogeneity (e.g. Goodchild, 2004). In the specific case of transportation systems, for example, many analysts have first-hand experience of

* Fax: +1 905 546 0463.
E-mail address: paezha@mcmaster.ca

0966-6923/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtrangeo.2005.11.002

geographical and socio-economic variations due to heterogeneity: tastes with regards to modal choice may not be the same at different levels of income (socio-economic context), and trip generation, distribution and commuting preferences often show clear geographical patterns (spatial context). Over the years, a number of approaches have been proposed to solve the problem of non-stationary relationships (e.g. market segmentation, dummy variables, and multilevel models). These approaches are satisfactory in many types of applications, in particular when contextual dimensions can be discretely classified (for example, by gender, ethnicity or employment type). However, these schemes may be of more limited appeal when there is a need to study contextual effects in a continuous-space setting, such as geographical space.

The objective of the present paper is to propose a simple binary probit model for continuous-space contextual effects. This line of inquiry is relevant to current research activities that explore the effects of space in transportation and land use modeling. The present paper follows directly from recent work in the field of spatial analysis, in particular local forms of spatial analysis (Fotheringham and Brunsdon, 1999). As documented by Páez and Scott (2004), it also attends to an incipient, but growing interest in transportation and land use studies concerning the implications and potential of spatial effects in model development and testing (see for example Bhat and Zhao, 2002; Bhat and Guo, 2004; Miyamoto et al., 2004; Páez and Suzuki, 2001).

The structure of the paper is as follows. In the following section, the relevant literature is briefly reviewed to place the paper within the context of existing methods. Next, the method of geographically weighted regression, a local form of spatial analysis on which the proposed model is based, is described and discussed. A heteroscedastic probit model with geographical weights is then proposed, and its use exemplified using land use data from California's BART system. Finally, some conclusions are drawn and directions for future research are sketched.

2. Background

A characteristic of many statistical methods is the implicit assumption that relationships are constant over the space of the sample; in other words, coefficients are assumed to be identical, or stationary, for all individuals, locations, zones, etc. within the sample. There are many situations, however, in which the assumption of stationarity is violated and/or difficult to maintain. For example, when studying the demand for public transportation, it is pertinent to ask whether the variables explain the response identically at different times of the day (temporal context), at different locations (geographical context) or for users with different income levels (socio-economic context). Over the years a number of methods have been developed to deal with these types of questions. These include:

(1) Market segmentation (e.g. Ben-Akiva and Lerman, 1985, pp. 202–204). This is a relatively simple procedure that can be applied to the data before a formal modeling effort, whenever the relationships in different regions or between socio-economic groups are believed to be non-constant. Segmentation means that the sample is subdivided into a small number of mutually exclusive and collectively exhaustive sub-samples (e.g. samples in geographical space, income, type of job, etc.). The method of market segmentation is simple to implement. Since it is based on standard discrete choice models, existing software can be used for estimation and testing. Further, it is amenable to hypothesis testing using, for example, likelihood ratio tests. The method is also limited and there are at least three problems with it: Firstly, it is not always clear how to divide the sample, and the classifications tend to be arbitrary. In a spatial context, this leads to the problem of modifiable areal units: the results of the model will be highly dependent upon the definition of the zoning system (Openshaw and Taylor, 1979). Secondly, the estimation scheme of separate models for each sub-sample means that the procedure does not make efficient use of total information content, since it effectively isolates each sub-sample (i.e. each market segment is taken out of context). As a secondary consequence of this, the impact of other market segments on the segment of interest is ignored. Finally, the segmentation scheme implicitly assumes that the sample can be divided into discrete categories. Although such a classification may be reasonable in many cases, it involves discontinuities that may not be reasonable when the contextual dimension is continuous. When areal zoning systems are used, individuals in different zones are assumed to be systematically different regardless of how short the distance between them is, while individuals who are far apart, but within the same zone will be identical from a spatial perspective.

(2) Dummy variables. The variables in this case relate to coefficients that are specific to a zone, socio-economic group, class, or individual. Dummy variables can be used to represent contextual effects, as for example in a model that classifies observations into two categories, say 'downtown' and 'suburbs', or 'high income', 'medium income' and 'low income'. Dummy variables can be tested by means of standard t-values, and their use helps to overcome the problem of non-linearities in relationships (see Ortúzar and Willumsen, 2001, pp. 139–140). However, as with the procedure of market segmentation, this form of segmentation involves discontinuities at the interface between classes or regions. The use of dummy variables loses further appeal when we note that the classifications are often arbitrary (an exception to

arbitrary classifications is the use of switching regressions with implicitly determined regimes, e.g., in population density analysis; see Alperovich and Deutsch, 2002).

(3) The expansion method (Casetti, 1972). A more sophisticated form of modeling contextual variations is by means of the expansion method. A variant of this has recently been proposed in the transportation literature as a flexible way to use socio-economic information to model individual tastes (Rizzi and Ortúzar, 2002). The method operates by expanding the coefficients of an initial model as functions of expansion variables, to obtain a terminal model that incorporates contextual effects. Although the expansion method does not implicitly involve discontinuities of the estimated parametric surface, it has been criticized because it is limited to deterministic expansions (Jones and Bullen, 1994). In addition, depending on the form of the expansion (i.e. the order of the polynomial, such as linear, quadratic, cubic, etc.), it might arguably fail to capture more complex variation patterns (Fotheringham et al., 1998). Higher order expansions may pose interpretability problems, as the interactions become increasingly complex.

(4) Multilevel models. This modeling approach has a relatively long history in geographical analysis (Jones, 1991), and in other disciplines where it is referred to variously as "hierarchical", "random-" or "varying-coefficients" models (e.g. Lawson et al., 2003). More recently, this modeling form has also attracted the attention of transportation/land use modelers (e.g. Bhat and Zhao, 2002). Multilevel models operate on a principle similar to that of the expansion method, but allow the introduction of random components as part of the expansion. Operationally, individuals in the sample can be nested in one or more higher levels according to districts, schools, age groups or any other suitable classification. The use of hierarchical levels and classifications is an attractive concept because heterogeneity in several dimensions can be simultaneously defined and estimated. Multi-level models, however, share some shortcomings with the methods previously described. In addition to the difficulty of having to categorize the observations prior to the analysis, an important consequence of introducing a random element in the expansion is that parametric variation becomes fragmented, giving rise to discontinuities in the boundaries between classes or regions. In addition, it could be argued that models of this type impose a hierarchical structure that may not be present in the process under analysis.

The objective of this paper is to propose a model that addresses some of the limitations of the approaches described above. To do this, we build upon the method of geographically weighted regression (GWR), a local form of spatial analysis useful to study geographical relationships (Brunsdon et al., 1996; Páez et al., 2002a). An important characteristic of GWR is its ability to estimate location-specific relationships that are tied to a focal (or regression) point. Local coefficient estimates can be obtained by displacing the focal point. This generates smooth parametric surfaces useful to study spatial stationarity or lack thereof. In the past, GWR has been developed mainly within a linear regression framework (but see the geographically weighted logistic regression developed by Atkinson et al., 2003). In order to broaden the range of applications of the method to situations commonly encountered in land use and transportation analysis, we extend some of these results to introduce continuous-space contextual effects into a model for limited-dependent variables. In the following section, GWR is briefly described and discussed.

3. Geographically weighted regression

Geographically weighted regression was proposed by Fotheringham, Brunsdon and Charlton (Brunsdon et al., 1996; Fotheringham et al., 1998) as a simple, but powerful method to study the issue of complex spatial parametric variation or spatial non-stationarity. The method, in essence a locally weighted regression, operates by assigning a weight to each and every observation i depending on its distance from a specific geographical location o termed the focal or regression point. The weighting system is based on the concept of distance-decay, made operational by means of a kernel function that reduces (i.e. down-weights) the influence of distant observations on estimation for location o (which could be any point including i), while implicitly emphasizing the influence of neighboring observations.

A more recent proposal uses distance-based variance functions to investigate non-stationarity (Páez et al., 2002a,b). In this version of the method, the geographical weights are applied to the variance of the error terms to give a form of heterogeneity that depends on relative location. To understand the operation mechanism of the method, begin by considering the classical regression model:

$$y_i = \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \tag{1}$$

where the constant term is obtained by defining $x_{i1} = 1$ for all i. It is usually assumed that the error terms ε_i in (1) are normally and independently distributed, with an expected value of 0. In addition, it is assumed that the variance of the error terms is constant:

$$E[\varepsilon_i^2] = \sigma^2 \quad \text{for all } i \tag{2}$$

Relaxing the assumption of constant variance leads to a broader class of models that reveals a number of interesting modeling possibilities. The general form of a non-constant variance model is defined by the following set of expressions (see Davidian and Carroll, 1987):

$$E[\varepsilon_i^2] = \varpi_{ii} = \sigma^2 g(\boldsymbol{\gamma}, \boldsymbol{z}_i) \tag{3}$$

$$E[\varepsilon_i \varepsilon_j] = 0 \tag{4}$$

In the above formulation the variance corresponding to error term i is modeled as a function of a $(p \times 1)$ vector of known variables $\boldsymbol{z}_i$ associated with vector $\boldsymbol{\gamma}$ $(p \times 1)$, and a base variance parameter σ^2. To ensure regularity conditions, it is assumed that $\omega_{ii} > 0$ for all i, and that ω_{ii} is a twice differentiable function of σ^2 and $\boldsymbol{\gamma}$. The alternative model of constant variance is obtained by assuming that there is a unique set of values $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ for which the function $g(\boldsymbol{\gamma}, \boldsymbol{z}_i) = 1$ for all i. This of course means that Eq. (3) reduces to Eq. (2) to give the classical regression model.

Different functions have been proposed to model the variance. For instance Davidian and Carroll (1987) consider, among other forms, a model of the variance that is quadratic in the explanatory variables ($p = 2K$):

$$g(\boldsymbol{\gamma}, \boldsymbol{z}_i) = 1 + \sum_{p=1}^{k} \gamma_p x_{ip} + \gamma_{p+k} x_{ip}^2 \tag{5}$$

The above equation suggests a spatial model for the variance that takes the form of a quadratic trend surface, if we let the known vector $\boldsymbol{z}_i$ (6×1) be the coordinates (c_x, c_y), squared coordinates, and c_x–c_y interactions of location i:

$$g(\boldsymbol{\gamma}, \boldsymbol{z}_i) = 1 + \gamma_1 c_{xi} + \gamma_2 c_{yi} + \gamma_3 c_{xi}^2 + \gamma_4 c_{xi} c_{yi} + \gamma_5 c_{yi}^2 \tag{6}$$

In this case $p = 5$ and the null hypothesis of variance homogeneity is given by $\boldsymbol{\gamma} = \boldsymbol{\gamma}^* = \mathbf{0}$. An alternative variance model takes the exponential form, a function that is parsimonious and has the desirable property of being strictly positive.

$$g(\boldsymbol{\gamma}, \boldsymbol{z}_i) = \exp(\boldsymbol{z}_i' \boldsymbol{\gamma}) \tag{7}$$

The above models with non-constant variance address the problem of variable dispersion in the distribution of the error terms, but do not address the question of parametric non-stationarity. There are a number of ways to define a non-constant variance model that is at the same time a model with variable coefficients (e.g. multi-level models). A different approach that does not imply discontinuities or within-class constant relationships is to adopt an exponential function for the variance, as in (7), using geographical distance from a given point as the explanatory variable. The following is an example of a distance-based variance function:

$$\varpi_{ii} = \sigma_o^2 \exp(\gamma_o d_{oi}^2) \tag{8}$$

In this case, the variance is defined as a function of two parameters, namely σ_o^2 and γ_o, and one explanatory variable, that is, the distance between the focal point o and observation i. The above specification complies with the usual regularity conditions: clearly, the ith diagonal element of the covariance matrix $\omega_{oi} > 0$ as long as $\sigma_o^2 > 0$ (the usual non-negativity condition of the variance), and ω_{oi} is a twice differentiable function of σ_o^2 and γ_o.

A useful property of the above function is that it is not translation invariant. This means that as the focal point is displaced, the estimated values of the coefficients vary as they attempt to approximate an underlying non-stationary parametric surface. As a consequence of this property, GWR can be defined as a locally linear model in the mean, with a location specific vector of coefficients $\boldsymbol{\beta}_o$ as follows:

$$y_o = \sum_{k=1}^{K} \beta_{ok} x_{ok} + \varepsilon_o \tag{9}$$

Again, defining $x_{o1} = 1$ at every point gives the intercept of the equation. Note that now the coefficients of the model are not spatial constants, but correspond to focal point o. In addition, all coefficients and diagnostics are assigned to this point, to give a self-contained model that applies to this location alone. It is important to note that a local model is completely defined by the $2 + K$ vector of parameters $\boldsymbol{\theta}_o = [\boldsymbol{\beta}_o', \sigma_o^2, \gamma_o]$ in the same way that a model with non-constant variance is complete. The only difference, in fact, is that the origin of the explanatory variable (i.e. distance) is relocated to conduct local estimation. The underlying model of homogeneity is given by $\gamma_o = \gamma_o^* = 0$, in which case the model reduces to the usual constant variance situation and what might be termed the *global model* since the variance does not depend on location, and consequently neither do other parameters. By allowing the possibility that the variance varies in a given geographical direction we define a model of heteroscedasticity that is simple and that has a clear spatial interpretation.

The power of adopting a specific geographical location for estimation is that it allows us to move from a global to a local perspective of the problem. In this sense, an important characteristic of GWR is that it is not limited to estimation at the location of recorded observations (i.e., location o needs not be one in the sample set), and in general any number of models can be estimated $(o = 1, \ldots, m)$, with $m \geqslant 1$ and no implied relationship between m and the number of observations n. The method can be seen as the experiment of observing the same dataset from a number of different perspectives—meaning that we enjoy the advantage of being able to study the parametric 'landscape' from different vantage points. In practice models can be estimated and tested on a location-specific basis, for a single location of interest, or for $m > 1$ when non-stationarity is of interest.

4. Heteroscedastic probit with geographical weights

Limited-dependent variable models, otherwise known as discrete choice models within a utility-maximization frame-

work, have a long history of development and applications in transportation planning, in particular in the field of travel demand modeling (McFadden, 1974; Yai, 1989). These models are characterized by the use of dependent variables that assume discrete, rather than continuous values. In a modal choice situation, for example, one and only one of a number of alternative modes is used. Likewise, when modeling locational choice, the decisions involved are discrete in nature: either a location is taken or it is not (Anas, 1982). Seen from the locator's perspective, only one location may be chosen: the outcome of the decision is limited. These two are examples of problems that interest transport and land use modelers. For a long time it was assumed that the use of locational variables (e.g. distance to the CBD) was enough to explain the geography of the problem; it is now realized that geography plays a more critical role than previously thought. For example, it has been argued, in the context of discrete choice models, that the process of making a decision might be influenced by the behavior and opinions of agents with whom there is contact (Case, 1992). Adopters of a new technology, for example, influence others in the 'neighborhood' with the result that some may be encouraged to become adopters themselves. This introduces the concept of *neighborhood* (locations that share some affinity such as contiguity, which is geographical in this case) and the necessity to consider neighborhood effects in the model. Definition of the neighborhood presents an additional challenge, as a neighborhood will probably be, in effect, the manifestation of complex contextual effects that traditional modeling approaches do not strive to capture.

Consider for example the case of land use change (or development). In addition to factors such as accessibility and availability of land, which can be and have been analyzed using standard limited-dependent variable models (Landis et al., 1995; McMillen, 1989), it is reasonable to expect that land developments of one kind will tend to encourage similar changes at adjacent locations. Landis et al. describe three reasons why we would expect this: reduced development costs (e.g. infrastructure and public services), the possible existence of agglomeration economies, and land-use regulations. In addition to this, there is a possibility that the explanatory variables will exert their influence in different ways at different locations. For example, accessibility to a large transportation facility may elicit a different response depending on location. Locating too close to the facility may result in undesirable effects such as congestion, pollution and noise. On the other hand, locating too far from the facility may reduce overall accessibility to employment centers, recreational and other activities. Availability of vacant land may encourage development in certain areas (for example in the urban–rural interface, as the cost of opportunity increases) but discourage change or development in other areas, where a surplus of vacant land may be seen as a liability instead of an asset. The interest then is to apply the concept of geographical weights described before, in order to obtain a spatial limited-dependent variable model for the analysis of contextual effects. Model development is described next.

Under the limited-dependent variable approach of the probit model, it is assumed that there is an underlying response variable defined in regression form by

$$y_i^* = \boldsymbol{X}_i\boldsymbol{\beta} + u_i \tag{10}$$

In the above expression $\boldsymbol{X}_i$ is a $1 \times k$ vector of characteristics or explanatory variables and $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients. It is assumed that, in practice, the response variable is unobservable (hence the sometimes used term of latent variable models), and instead what is observed is a dummy variable defined by

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The dummy variable determines which of two possible outcomes is observed. From the above, it follows that the probability of observing outcome 1, when $y_i = 1$, is

$$\begin{aligned} \Pr(y_i = 1) = \Pr(y_i^* > 0) &= \Pr\left(u_i > -\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma_i}\right) \\ &= 1 - \Phi\left(-\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma_i}\right) \end{aligned} \tag{12}$$

whereas the probability of observing outcome 2 is

$$\Pr(y_i = 0) = 1 - \Pr(y_i = 1) \tag{13}$$

In the above expressions, Φ is a cumulative distribution function for error term u_i. Adoption of the logistic distribution leads to the logit model. The cumulative normal distribution (with σ_i as a variance or scale parameter), on the other hand, leads to the probit model with its more flexible covariance structure. A common simplifying assumption for the above model is that the variance is constant. When this assumption is accepted, the value of σ_i can be, and usually is arbitrarily determined. The probabilities of observing either outcome do not change when the coefficients are rescaled by a positive constant. It is thus usually assumed that $\sigma_i = \sigma = 1$ (Ben-Akiva and Lerman, 1985). The assumption of constant variance, or homoscedasticity, while the most common approach, can make estimation of the coefficients inconsistent (McMillen, 1992). In addition, it could mask potentially interesting variation across the parametric landscape.

In order to introduce contextual effects, which in this case are a consequence of heterogeneity, the variance parameter is defined by adopting a distance-based variance function such as discussed in the previous section. Thus, instead of assuming a constant and arbitrary value for σ_i, this parameter is given by

$$\sigma_i = g(\gamma_o, d_{oi}) = \exp(\gamma_o d_{oi}^2) \tag{14}$$

where the parameters are defined as before.

The log-likelihood of the above model is given by the following expression:

$$\ln L = \sum_{i=1}^{n} y_i \ln\left[\Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\exp(\gamma_o d_{oi}^2)}\right)\right] + \sum_{i=1}^{n}(1-y_i)\ln\left[1-\Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\exp(\gamma_o d_{oi}^2)}\right)\right] \quad (15)$$

The model has a total of $k+1$ locally linear parameters, in vector form $\boldsymbol{\theta}_o = [\boldsymbol{\beta}_o', \gamma_o]$. Note that the variance takes a value of 1 when $d_{oi} = 0$, that is, at the focal point. The variance at other locations will be given relative to the variance at the focal point. To obtain estimates of these parameters, the log-likelihood function is maximized with respect to them. McMillen (1992) presents the following procedure to estimate heteroscedastic probit models.

Define the following terms for the standard normal distribution and the cumulative normal distribution functions evaluated at point i:

$$\phi_i = \phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma_i}\right) \quad (16)$$

$$\Phi_i = \Phi\left(\frac{\boldsymbol{X}_i\boldsymbol{\beta}}{\sigma_i}\right) \quad (17)$$

Also, define a matrix $\mathbf{R}$ $[n\times(k+1)]$ as follows:

$$\mathbf{R} = \begin{bmatrix} \frac{\phi_1 X_{11}}{\sigma_1} & \cdots & \frac{\phi_1 X_{1k}}{\sigma_1} & -\frac{\phi_1 \boldsymbol{X}_1\boldsymbol{\beta}}{\sigma_1} d_{o1}^2 \\ \frac{\phi_2 X_{21}}{\sigma_2} & \cdots & \frac{\phi_2 X_{2k}}{\sigma_2} & -\frac{\phi_2 \boldsymbol{X}_2\boldsymbol{\beta}}{\sigma_2} d_{o2}^2 \\ \vdots & & \vdots & \\ \frac{\phi_n X_{n1}}{\sigma_n} & & \frac{\phi_n X_{nk}}{\sigma_n} & -\frac{\phi_n \boldsymbol{X}_2\boldsymbol{\beta}}{\sigma_n} d_{on}^2 \end{bmatrix} \quad (18)$$

a diagonal matrix $\mathbf{H}$ $(n\times n)$:

$$\mathbf{H} = \begin{bmatrix} \frac{1}{[\Phi_1(1-\Phi_1)]^{1/2}} & & & 0 \\ & \frac{1}{[\Phi_2(1-\Phi_2)]^{1/2}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{[\Phi_n(1-\Phi_n)]^{1/2}} \end{bmatrix} \quad (19)$$

and a vector $\boldsymbol{e}$ $(n\times 1)$ as follows:

$$\boldsymbol{e} = \begin{bmatrix} y_1 - \Phi_1 \\ y_2 - \Phi_2 \\ \vdots \\ y_n - \Phi_n \end{bmatrix} \quad (20)$$

The vector of first derivatives can then be expressed as

$$\frac{\partial}{\partial\boldsymbol{\theta}_o}\ln L = \mathbf{R}'\mathbf{H}'\mathbf{H}\boldsymbol{e} \quad (21)$$

while the expected value of the second derivative matrix is

$$E\left[\frac{\partial^2 \ln L}{\partial\boldsymbol{\theta}_o\partial\boldsymbol{\theta}_o'}\right] = -\mathbf{R}'\mathbf{H}'\mathbf{H}\mathbf{R} \quad (22)$$

Estimation follows an iterative process, with the change in estimates between steps given by

$$\boldsymbol{\theta}_o^s - \boldsymbol{\theta}_o^{s-1} = (\mathbf{R}'\mathbf{H}'\mathbf{H}\mathbf{R})^{-1}\mathbf{R}'\mathbf{H}'\mathbf{H}\boldsymbol{e} \quad (23)$$

As noted by McMillen (1992) the estimator is equivalent to a non-linear weighted least squares estimator, derived from the regression of $\boldsymbol{e}$ on $\mathbf{R}$ with weights given by matrix $\mathbf{H}$. Initial values for the optimization procedure can be obtained if a regular (i.e. homoscedastic) probit model is estimated.

5. Example: Transportation infrastructure and land use change

5.1. Study area and data

The example in this paper reanalyzes data collected and analyzed by Landis et al. (1995) from California's Bay Area Rapid Transit (BART) system. One objective of the analysis by Landis and colleagues was to assess the impact of transit investment (i.e. transportation infrastructure) on land uses, and in particular whether or not proximity to a BART station encouraged new development or redevelopment of existing areas. Several stations were studied, and a regular grid was superimposed to produce square zones of 1 ha in which land uses where measured. A number of variables were examined to explore their effect on land use change. These included a measure of proximity to a BART station (which was not found to be a significant explanatory factor in any of the stations studied), and a measure of intervening opportunities in the form of percentage of land available for new development closer to the station (also not significant). In addition to these variables, a measure of interaction or cohesiveness among locations was introduced in the form of a 'similarity index', used to describe the prevailing land use in the eight immediately adjacent cells to each location. In most cases, Landis et al. (1995) found that high similarity indices (i.e. similar land use in the neighborhood) tended to discourage further land use change. Similar findings were obtained in a reanalysis of the data, in an example that applied spatial modeling techniques to explore the issue of the dynamics of residential land use change (Páez and Suzuki, 2001).

In this section, we reanalyze the data from Union City BART station using the probit model with geographical weights introduced in the preceding section. This particular station is an example of a location where the present pattern of land uses was determined between 1965 and 1990, concurrently with the development of BART. Unlike previous studies by Landis et al. (1995) and Páez and Suzuki (2001), the objective here is not to incorporate spatial interactions between locations (i.e., similarity or autocorrelation), but to explore the issue of coefficient variability in geographical space. In this sense, the analysis looks at a different form of spatial effects. The dependent variable in the analysis that follows indicates land use change by use of

two possible values: 1 if land use changed and 0 otherwise. Two variables from the original analysis are used: distance to station (station is at the center of the area), and percentage of undeveloped land use closer to the station. In total there are 324 cells in the study region, covering an area of 3.24 square kilometers around Union City BART Station. Of the 324 cells in the sample, 61 changed to residential land use in the period between 1965 and 1990.

5.2. Results and discussion

The first step in this empirical section was to estimate a regular binary probit model with constant variance, to serve as a benchmark model to compare the results of the local modeling exercise. The results of estimating such a model appear in Table 1. The variables used are as described in the preceding section. The table shows that, apart from the constant, the coefficients of the model are not significantly different from zero. These results confirm previous findings by Landis et al. (1995), who used a binary logit model and a different combination of variables, and Páez and Suzuki (2001), who used a dynamic spatial logit model. These studies suggest that proximity to transit service did not contribute substantially to determine land use change in this case.

As a second step, a set of local coefficients was estimated using the proposed probit model with distance-based variance functions. An attempt was made to obtain local coefficient estimates at each of the 324 cells in the study area. Of these, the iterative process used to estimate the coefficients did not converge in a number of locations (29 cells, or about 9% of the points in the sample). Upon closer examination, it was found that the points where estimation failed combined values of the explanatory variables and parameters in the variance that lead to undefined values of the log-likelihood function. More generally, when there is local zero variance in the response variable, or when the variance is extremely high, the model cannot be estimated for the location [see Eq. (15)]. A similar problem could be expected in cases when there is strong collinearity among some or all explanatory variables, which would require a reassessment of their selection. In the present case it is possible to represent the landscape of general variation in the study area by interpolating the coefficient values at ill-conditioned locations using the values at surrounding locations. Interpolation is commonly used when preparing maps of local coefficient variability with linear regression (see for example Fotheringham et al., 2002), and appears to be justified in the case of the model here by the fact that the spatial distribution of the coefficient estimates supports the idea of spatially continuous variation before interpolation. As in the case of conventional GWR models, it is important to note that the objective of interpolation is for presentation only, and that estimation results are strictly valid only for those locations where coefficients and diagnostics are actually obtained. The results appear in Figs. 1–3.

Regarding the spatial variation of the coefficients, the results for the constant suggest that land use change is less likely to take place in the immediate surroundings of the station; a clear west–east trend suggest that land use change is a more likely outcome to the east of the station. This east–west trend is observed also in the variation of coefficients β_2 and β_3. An interesting exercise is to calculate

Table 1
Logit model

Variable	Coefficient	Estimate	Standard deviation
Constant	β_1	−3.011	0.936
Distance to station	β_2	3.287	2.64
% of vacant land closer to station	β_3	−1.002	2.189

Log-likelihood = −121.98.

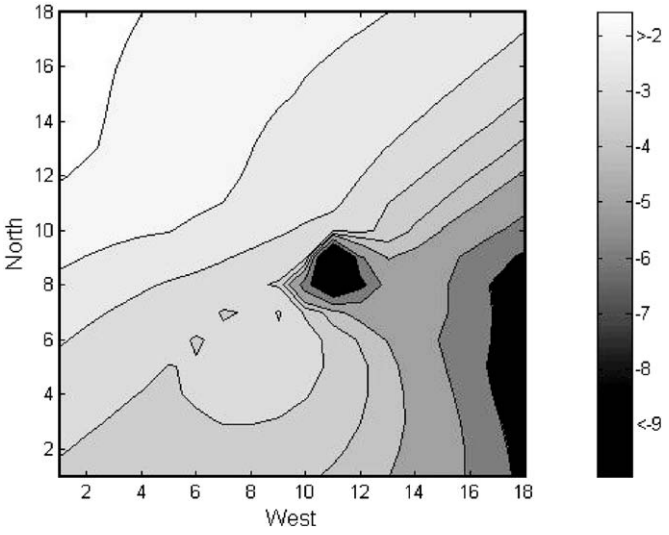

Fig. 1. Spatial distribution of coefficient β_1 (constant).

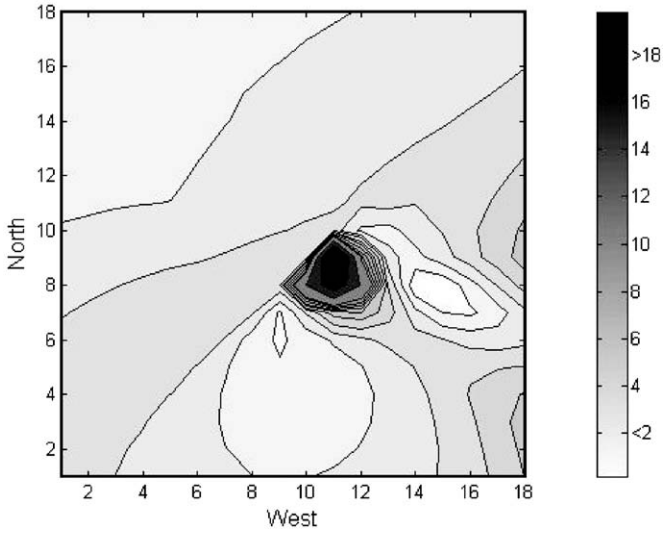

Fig. 2. Spatial distribution of coefficient β_2 (distance to station).

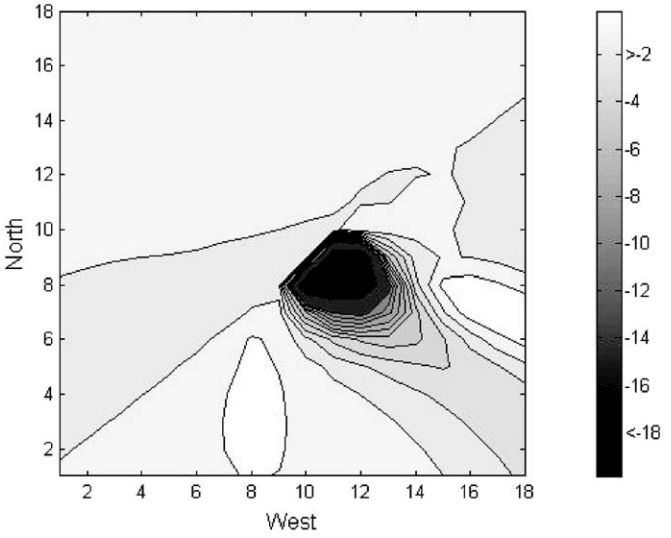

Fig. 3. Spatial distribution of coefficient β_3 (% of vacant land closer to station).

the *t*-scores of the local coefficients. The constant, which was significant in the global model, was also significant for most of the region in the local modeling approach. Still, it was found that for at least 25 locations in the study area, this coefficient was not significant. On the other hand, the coefficient associated with distance from the station, which was found to be not significant in the global model, turned out to be significant for at least 40 cells in the area surrounding the station in the local modeling approach. This result suggests that while the presence of the station may not have had an effect on land use change at longer distances, it may have decreased the likelihood of change to residential uses at short distances (i.e. in areas near the station).

The goodness of fit of the local estimations was evaluated by means of the value of the log-likelihood function. It was found that the local log-likelihood ranges between −121.97 and −100.48. Furthermore, the likelihood ratio tests (χ^2 distributed with 1 degree of freedom, reflecting one constraint being imposed on the restricted model) show that the local model produces a higher likelihood in as many as 215 locations when compared to the model with constant variance and global coefficients. Although the test was not adjusted for multiple comparisons (this would require an adjustment to the nominal level of significance based on the Bonferroni or a similar approach; see Páez et al., 2002a), it is still clear that the local modeling approach can improve the performance of the model.

5.3. Land use change: relation to other analytical approaches

The lattice used for the analysis of land use change in the example above is reminiscent of the lattices used in cellular automata (CA) simulations. Cellular automata are models in which each cell in a (usually regular) lattice may be in one of a series of defined states (defined by the attributes or characteristics of the cell, i.e., alive or dead, developed or undeveloped, etc.) Moreover, the state of a cell may change based on the repetitive application of simple transition rules that depend on the situation in the neighborhood, or in other words, on the status of neighboring/contiguous/nearby cells (Batty, 1997). As noted by Batty (1997), transition rules in CA can be interpreted as generators of urban growth or decline—in essence land use change. CA models have attracted substantial attention as a tool to model self-organizing cities, urban form, and complex systems (for recent work see Caruso et al., 2005; Wu and Webster, 2000). Application of the probit model with geographical weights in the example above shares with CA the following characteristics (see Torrens and O'Sullivan, 2001): the previously noted spatial lattice, the set of allowed states for the cells (i.e. residential and non-residential in the example), and the spatial effect determined by the neighborhood. The probit model itself could be seen as the embodiment of a probabilistic transition rule. On the other hand, the probit model, being a cross-sectional model, lacks the dynamic aspects of CA. Despite this difference, the model in this paper, as well as the dynamic spatial logit model developed by Dubin (1995), and used by Páez and Suzuki (2001) to study land use change, could be seen as a complement to CA models. While the dynamic spatial logit model explicitly incorporates the status of cells in the neighborhood in the transition probabilities, the model and example in this paper suggest that transition rules may not be all that simple, and that they could, in fact, vary by location. Indeed, statistical and econometric models such as those mentioned here could be used as tools to investigate transition rules, and to validate the process of change in CA. Process validation is an area currently underdeveloped in CA research, with much of existing work concentrating on validation of form by means of pattern recognition approaches (Torrens and O'Sullivan, 2001).

6. Conclusions

Traditional statistical methods commonly used to study land use and transportation systems tend to ignore or underestimate the importance of geography. This is now beginning to change, as more research is devoted to the study of spatial effects in models for land use and transportation analysis. This paper aims at contributing to this relatively recent strand in the literature. The objective of the paper has been to propose a simple binary probit model with geographical weights, useful to explore the issue of spatial parametric non-stationarity, that is, the variation of coefficients in geographical space. The method follows on the heels of recent developments in the geographical analysis literature, in particular heteroscedastic probit models (McMillen, 1992), and the method of GWR (Brunsdon et al., 1996; Páez et al., 2002a). The model proposed was applied to a case study that assesses the impacts of transportation infrastructure on land use change. The

specification of the model in the example was very simple, with only two variables, but sufficient to demonstrate the relevance of exploring spatially varying relationships. Two findings are worth noting. The first relates the degree to which relationships vary in the study area. The level of variation suggests that non-stationarity should be seriously considered, as it may mask potentially large differences in the operation of the process. Coefficients that were not significant in the global model, for example, were significant at some locations when local models where estimated. The second finding relates to the statistical fit of the local models, which was found, by means of a likelihood ratio test, to be higher than that of the global (homoscedastic) model.

An important issue detected in the application of the model, on the other hand, was its inability to converge at certain locations. Detailed exploration of these locations showed that certain combinations of explanatory variable values and coefficients estimates (for example a negative value of γ_o with a large distance value in the variance component of the specification), can cause problems in estimation. In cases like these, very small probability values lead to indeterminate values of the log-likelihood function. Although a way around this issue is to interpolate using estimated values, this remains a feature of the model that requires attention in future research. In relation to this, the issue of collinearity and/or local zero variance should be further explored. Possible avenues of research include the adoption of semi-parametric methods, in which the geographical weight parameter is exogenous (and bounded). In addition, it would be interesting to combine the proposed model for heterogeneity in continuous space with some recent work that incorporates spatial autocorrelation in discrete choice models (e.g. Bhat and Guo, 2004; Miyamoto et al., 2004; Mohammadian and Kanaroglou, 2003).

Acknowledgement

The research reported in this paper was supported by Canada's Natural Sciences and Engineering Research Council (NSERC) grant 261782-03. The author is grateful to Mr. Patrick DeLuca, Prof. Bruce Ralston, and the anonymous reviewers who provided valuable comments on a previous version of this paper.

References

Alperovich, G., Deutsch, J., 2002. An application of a switching regimes regression to the study of urban structure. Papers in Regional Science 81 (1), 83–98.

Anas, A., 1982. Residential Location Markets and Urban Transportation. Academic Press, New York.

Atkinson, P.M., German, S.E., Sear, D.A., Clark, M.J., 2003. Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. Geographical Analysis 35 (1), 58–82.

Batty, M., 1997. Cellular automata and urban form: a primer. Journal of the American Planning Association 63 (2), 266–274.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Applications to Travel Demand. The MIT Press, Cambridge.

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transportation Research Part B—Methodological 38 (2), 147–168.

Bhat, C., Zhao, H.M., 2002. The spatial analysis of activity stop generation. Transportation Research Part B—Methodological 36 (6), 557–575.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Analysis 28 (4), 281–298.

Caruso, G., Rounsevell, M., Cojocaru, G., 2005. Exploring a spatio-dynamic neighbourhood-based model of residential behaviour in the Brussels periurban area. International Journal of Geographical Information Science 19 (2), 103–123.

Case, A., 1992. Neighborhood influence and technological-change. Regional Science and Urban Economics 22 (3), 491–508.

Casetti, E., 1972. Generating models by the expansion method: applications to geographic research. Geographical Analysis 28, 281–298.

Davidian, M., Carroll, R.J., 1987. Variance function estimation. Journal of the American Statistical Association 82, 1079–1091.

Dubin, R., 1995. Estimating logit models with spatial dependence. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 229–242.

Fotheringham, A.S., Brunsdon, C., 1999. Local forms of spatial analysis. Geographical Analysis 31 (4), 340–358.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester.

Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30 (11), 1905–1927.

Goodchild, M.F., 2004. The validity and usefulness of laws in geographic information science and geography. Annals of the Association of American Geographers 94 (2), 300–303.

Jones, K., 1991. Specifying and estimating multilevel models for geographical research. Transactions of the Institute of British Geographers 16 (2), 148–159.

Jones, K., Bullen, N., 1994. Contextual models of urban house prices—a comparison of fixed-coefficient and random-coefficient models developed by expansion. Economic Geography 70 (3), 252–272.

Landis, J., Guhathakurta, S., Huang, W., Zhang, M., Fukuji, B., Sen, S., 1995. Rail transit investments, real estate values, and land use change: a comparative analysis of five California Rail Transit Systems. Institute of Urban and Regional Development, University of California at Berkley.

Lawson, A.B., Brown, W.J., Vidal-Rodeiro, C., 2003. Disease Mapping using WinBUGS and MLwIN. Wiley, London.

Martínez, F.J., 1992. The bid-choice land use model: an integrated economic framework. Environment and Planning A 24, 871–885.

McFadden, D., 1974. The measurement of urban travel demand. Journal of Public Economics 3, 303–328.

McMillen, D.P., 1989. An empirical-model of urban fringe land-use. Land Economics 65 (2), 138–145.

McMillen, D.P., 1992. Probit with spatial autocorrelation. Journal of Regional Science 32 (3), 335–348.

McNally, M., 2000a. The activity-based approach. In: Hensher, D.A., Button, K.J. (Eds.), Handbook of Transport Modelling. Pergamon, Oxford, pp. 53–68.

McNally, M., 2000b. The four-step model. In: Hensher, D.A., Button, K.J. (Eds.), Handbook of Transport Modelling. Pergamon, Oxford, pp. 33–52.

Miyamoto, K. 1993. Development and applications of a land-use model based on random utility/rent bidding (RURBAN). In: 8th World Conference on Transport Research, Antwerp.

Miyamoto, K., Vichiensan, V., Shimomura, N., Páez, A., 2004. Discrete choice model with structuralized spatial effects for location analysis. Transportation Research Record 1898, 183–190.

Mohammadian, A., Kanaroglou, P.S., 2003. Applications of spatial multinomial logit model to transportation planning. In: Proceedings of the 10th International Conference on Travel Behaviour Research, Switzerland.

Openshaw, S., Taylor, P.J., 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley, N. (Ed.), Statistical Applications in the Spatial Sciences. Pion, London, pp. 127–144.

Ortúzar, J.D., Willumsen, L.G., 2001. Modelling Transport, third ed. Wiley, New York.

Páez, A., Scott, D.M., 2004. Spatial statistics for urban analysis: a review of techniques with examples. GeoJournal 61 (4), 53–67.

Páez, A., Suzuki, J., 2001. Transportation impacts on land use change: an assessment considering neighborhood effects. Journal of the Eastern Asia Society for Transportation Studies 4 (6), 47–59.

Páez, A., Uchida, T., Miyamoto, K., 2002a. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. Environment and Planning A 34 (4), 733–754.

Páez, A., Uchida, T., Miyamoto, K., 2002b. A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. Environment and Planning A 34 (5), 883–904.

Rizzi, L.L., Ortúzar, J.d.D., 2002. Stated preferences in the evaluation of interurban road safety. Accident Analysis and Prevention 35, 9–22.

Torrens, P.M., O'Sullivan, D., 2001. Cellular automata and urban simulation: where do we go from here? Environment and Planning B—Planning & Design 28 (2), 163–168.

Wu, F.L., Webster, C.J., 2000. Simulating artificial cities in a GIS environment: urban growth under alternative regulation regimes. International Journal of Geographical Information Science 14 (7), 625–648.

Yai, T., 1989. Disaggregate behavioural models and their applications in Japan. Transportation Research Part A 23 (1), 45–51.