Spatial nonstationarity and autoregressive models

C Brunsdon
Department of Town and Country Planning, University of Newcastle, Newcastle upon Tyne
NE1 7RU, England; e-mail: chris.brunsdon@newcastle.ac.uk
A S Fotheringham, M Charlton
Department of Geography, University of Newcastle, Newcastle upon Tyne NE1 7RU, England;
e-mail: stewart.fotheringham@newcastle.ac.uk; martin.charlton@newcastle.ac.uk
Received 19 August 1996; in revised form 21 February 1997

Abstract. Until relatively recently, the emphasis of spatial analysis was on the investigation of global models and global processes. Recent research, however, has tended to explore exceptions to general processes, and techniques have been developed which have as their focus the investigation of spatial variations in local relationships. One of these techniques, known as geographically weighted regression (GWR), developed by the authors is used here to investigate spatial variations in spatial association. The particular framework in which spatial association is examined here is the spatial autoregressive model of Ord, although the technique can easily be applied to any form of spatial autocorrelation measurement. The conceptual and theoretical foundations of GWR applied to the Ord model are followed by an empirical example which uses data on owner-occupation in the housing market of Tyne and Wear in northeast England where the problems of relying on global models of spatial association are demonstrated. This empirical investigation of spatial variations in spatial autocorrelation prompts a further discussion of several issues concerning the statistical technique.

Introduction

Over the last two or three decades, quantitative geographers and statisticians have tried to model the effects of spatial association in regression analysis. A well-known example of this effort is that by Ord (1975) who has proposed the use of *autoregressive* and *moving average* terms in regression models to account for spatial correspondences in the response variable and the residuals, respectively. Although we appreciate the contribution that this technique has made to the consideration of spatial process in regression modelling, it, along with similar efforts to account for spatial autocorrelation, can be criticised on the grounds of producing global results which assume that spatial processes operate uniformly throughout the study area. There has been a recent shift in emphasis in spatial statistics away from such 'global' types of analysis to 'local' ones where the aim is to identify spatial variations in relationships (Fotheringham, 1997). One of the earliest attempts to model local relationships is that of the expansion method (Casetti, 1972; Fotheringham and Pitts, 1995; Jones and Casetti, 1992) although, as demonstrated by Fotheringham et al (1997a), the expansion method essentially fits trends to surfaces of local relationships and can therefore miss important local variations which run counter to these general trends. Other examples of local statistical analyses include those of Anselin (1995), Brunsdon et al (1996), Fotheringham et al (1997b; 1997c), Getis and Ord (1992), and Ord and Getis (1995).

It is our aim in this paper to present a methodology that addresses the need for localised versions of spatially autoregressive models by producing, as an example of this type of modelling a localised version of Ord's model in which the output appears as a spatial distribution of localised values indicating local autocorrelation rather than a single global estimate. It should be noted that the method can be easily extended to

any framework in which local dependencies are measured and that the Ord model considered here is just one example of this type of analysis.

In Ord's original autoregressive model, the regression equation can be expressed as

$$y_i = \rho \sum_j w_{ij} y_j + e_i, \tag{1}$$

where y_i is an observation at point i, w_{ij} is the ijth element of $\mathbf{W}$, a contiguity matrix representing the spatial arrangement of a set of zones for which the ys are attributes, ρ is a coefficient of spatial cohesion, and e_i is a normally distributed error. It is clear that in this model the ρ term is fixed for all geographical locations. This implies that the level of spatial association between all adjacent zones is fixed. To take an example of the spatial distribution of house prices, the assumed form of equation (1) implies that the degree of influence of house prices in surrounding areas is the same every-where. An alternative conjecture might be that in some areas this spatial influence is more marked than in others. To address this possibility we propose a modified model of the form

$$y_i = \rho_i \sum_j w_{ij} y_j + e_i, \tag{2}$$

where i indexes a location in geographical space where data are observed.

Below, an overview of Ord's method is given, together with a slight modification of the model followed by a discussion of how this may be extended to a method for calibrating models such as equation (2). This is then extended to models in which predictor variables are added to the modelling framework. Finally, an example involving the spatial distribution of owner-occupied housing in the metropolitan county of Tyne and Wear in the United Kingdom from the 1991 population census is given.

Specifying the autoregressive model
A global autoregressive model, such as equation (1), encapsulates to some extent a spatial diffusion process affecting the y-variable. In the case of equation (1), it is assumed that each y-variable can be modelled in terms of a spatial smoothing of its neighbours, $\sum_j w_{ij} y_j$, and an error term e_i. If $\mathbf{W}$ is such that each of its rows sum to unity, and if the distribution of each element of y, the vector of the y_i, has a mean μ, then each element of the smoothed vector $\mathbf{W}y$ will also have the same mean value. This suggests that in a model such as equation (1) we have the following relationship between expectation values, E,

$$\mathrm{E}(y) = \rho\,\mathrm{E}(y) + \mathrm{E}(e), \tag{3}$$

and by rearrangement this gives

$$\mathrm{E}(y) = \frac{\mathrm{E}(e)}{1 - \rho}. \tag{4}$$

This contrasts with the more usual situation where the mean of the error term is zero, unless the y-variate itself has a mean of zero. A more satisfactory model might be

$$y_i - \mu = \rho \sum_j w_{ij} (y_i - \mu) + e_i. \tag{5}$$

In this case it is the deviation about a mean level that is assumed to be autocorrelated. This also gives a more reasonable meaning to the hypothesis $\mu = 0$ because this implies that the y-variate is independently distributed about a fixed mean value μ.

The model in equation (5) can be adapted to a model of varying spatial cohesion in the same manner as equation (1) and can be transformed into a form equivalent to equation (2):

$$y_i - \mu = \rho_i \sum_j w_{ij}(y_i - \mu) + e_i. \tag{6}$$

In this paper, models (2) and (6) will both be considered in order to investigate the degree of advantage gained by adopting equation (6) instead of equation (2).

Calibrating the autoregressive model

Clearly, in order to explore the geographical structure of spatial cohesion for a given data set, it is necessary to consider methods of calibrating equations (6) and (2). As will be seen below, approaching this task is best done by first considering the calibration of equations (5) and (1) for a given set of data. First, consider the simpler model (1). In vector notation this may be rearranged to give

$$y = (I - \rho W)^{-1} e,$$

or

$$Ay = e. \tag{7}$$

If we can assume that $E(y) = 0$, then it may be seen that y will be multivariate normal with a zero vector mean and a variance – covariance matrix of $(A^T A)^{-1}$. From Ord (1975) and Mead (1967), the likelihood function for an observed set of y-values, y, would be given by

$$l(\sigma^2, \rho) = \text{constant} - \frac{n}{2} \ln(\hat{\sigma}^2 |A|^{-2/n}), \tag{8}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2, $\hat{\sigma}^2 = y^T A^T A y / n$.

Because the scale parameter expression may be substituted into the likelihood function, the only remaining problem is the estimation of ρ. This is a problem of univariate optimisation for which there exist a number of numerical solutions. Ord, for example, suggests the use of Newton's method. Another possibility might be a golden section search (see, inter alia, Greig, 1980). In either case, the problem is equivalent to finding ρ which minimises

$$-\frac{2}{n} \ln|A| + \ln(e^T e). \tag{9}$$

Both terms in equation (9) depend on ρ—the first is a function of the determinant of A in equation (7) and the second is the logarithm of the residual sum of squares of model (1) for a given ρ. For further discussion of this expression of the likelihood function, see Anselin (1988).

Minimising equation (9) is computationally intense because of the first term, in which a determinant is computed. Ignoring this term is equivalent to selecting ρ on the basis of a least squares criterion. Again, Ord considers this option but finds that unless ρ is close to zero this leads to a notable degree of bias in the estimation. However, noting that

$$|A| = |I - \rho W| = \prod_{i=1}^{n}(1 - \lambda_i \rho), \tag{10}$$

where λ_i is the ith eigenvalue of W, we may express equation (9) as

$$-\frac{2}{n} \sum_{i=1}^{n} \ln(1 - \rho \lambda_i) + \ln(e^T e), \tag{11}$$

thus avoiding the computation of a determinant. The eigenvalues of $\mathbf{W}$ need only be computed once. Thus, the value of ρ minimising equation (11) is a maximum likelihood estimate from which the scale parameter may be derived by using equation (8).

It then remains to be seen how this technique may be adapted for a model in which ρ is not fixed over space but is allowed to vary with i. One approach would be to use a geographical weighting technique to estimate ρ_i as described in Brunsdon et al (1996) and Fotheringham et al (1997a; 1997b; 1997c). In such a technique, a weighted estimate of ρ_i is obtained for each point i by using data weighted according to location around point i and the weights being a monotone decreasing function of distance from i. For each point at which ρ_i is estimated, different sets of weights will therefore apply according to the location of i.

One way of applying a weighting scheme to model (1) is to allow the errors—the elements of the e vector—to have different variances. In maximum likelihood estimation the influence of zones for which a large error variance exists will be downweighted. Thus, in vector notation, model (1) now becomes

$$y = \rho_i\mathbf{W}y + \mathbf{D}_i e, \tag{12}$$

where $\mathbf{D}_i$ is a diagonal matrix of standard deviations of error terms which is used mainly as a device to introduce a weighting scheme emphasising observations around i. For zones near to the sample point i the standard deviations are low (and corresponding weights are large) whereas for zones further from i the standard deviations are high (and corresponding weights are small). This may then be rearranged so that

$$\mathbf{K}_i\mathbf{A}_i y = e, \tag{13}$$

where $\mathbf{K}_i\mathbf{D}_i = \mathbf{I}$. The maximum likelihood expression for equation (13) is simply equation (8) with $\mathbf{K}_i\mathbf{A}$ substituted for $\mathbf{A}$. Noting that the determinant of $\mathbf{K}_i\mathbf{A}$ is the product of the determinants of $\mathbf{K}_i$ and $\mathbf{A}$, and that the determinant of $\mathbf{K}_i$ is constant for any given ρ_i to be estimated, we find that a maximum likelihood estimator will minimise the expression

$$-\frac{2}{n}\sum_{j=1}^{n}\ln(1 - \rho\lambda_j) + \ln(e^{\mathrm{T}}e), \tag{14}$$

where λ_j is the jth eigenvalue of $\mathbf{K}_i\mathbf{A}$. Thus if we make $\mathbf{K}_i$ a function of distances between pairs of points, on the basis of a distance-decay function, n different weighting schemes will be generated with each $\mathbf{K}_i$ being used to estimate a corresponding ρ_i. Note that $\mathbf{K}_i$ is essentially a matrix representing a spatial kernel function. The jth diagonal element of $\mathbf{K}_i$ is a kernel function of the distance between the centroid of the jth zone and the point i at which a local value of ρ_i is to be estimated. This process will be relatively costly in computational terms because each ρ_i must be estimated iteratively but the eigenvector 'trick' for evaluation of determinants reduces the number of operations required considerably. Note also that the eigenvalues used are those of $\mathbf{W}$ and do not vary with location (unlike the values of $\mathbf{K}_i$) and need only be computed once at the outset of the analysis.

Next, the method may be extended from models of the form (2) to those of the form (6). Again if we look to Ord (1975), explanatory variables may be incorporated into an autoregressive model to give a new model of the form

$$y = \rho\mathbf{W}y + \mathbf{X}\beta + e. \tag{15}$$

Model (5) is a special case of this in which $\mathbf{X}$ is simply a column of ones. Maximum likelihood estimates of ρ may be obtained in a similar manner to those in model (1)

by finding ρ which minimises the expression

$$-\frac{2}{n} \sum_{i=1}^{n} \ln(1 - \rho\lambda_i) + \ln(e^{\mathrm{T}}e) . \tag{16}$$

It is then possible to estimate β by using

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \rho\mathbf{W})y , \tag{17}$$

where estimates of β and ρ are both maximum likelihood estimates.

As before, the next stage is to extend these models so that ρ is allowed to vary spatially. As with equation (12), this can be done by assuming a distance-based multiplier on the error variance, giving

$$y = \rho_i \mathbf{W}y + \mathbf{X}\beta_i + \mathbf{D}_i e , \tag{18}$$

which can be expressed as

$$\mathbf{K}_i\mathbf{A}y - \mathbf{K}_i\mathbf{X}\beta_i = e , \tag{19}$$

where $\mathbf{K}_i\mathbf{D}_i = \mathbf{I}$.

As before, it is then possible to find the maximum likelihood estimates of β_i and ρ_i by substituting $\mathbf{K}_i\mathbf{A}$ for $\mathbf{A}$ and $\mathbf{K}_i\mathbf{X}$ for $\mathbf{X}$. Again, this is equivalent to applying a spatial kernel multiplier to the rows of $\mathbf{A}$ and $\mathbf{X}$, centred around the point at which the parameters are to be estimated.

To apply a model of the type (6), $\mathbf{X}$ will simply be a column of ones, as suggested earlier. In this case, β_i is a scalar which varies over space so that the model is equivalent to a trend surface plus a spatially smoothed variation about this surface in which the degree of smoothing may vary locally.

An example based on owner-occupation housing data

To demonstrate the operation of the localised autoregressive models shown above, the spatial distribution of the percentage of home owner-occupiers in Tyne and Wear, a metropolitan county in the United Kingdom, is examined. Over the past two decades, there have been notable increases in the United Kingdom in the number of people buying homes, as opposed to renting. There are several explanations for this phenomenon, particularly the UK government's 'right-to-buy' initiative in which occupiers of council-owned housing were encouraged to buy their homes at subsidised rates.

Although some may argue that it has been advantageous to purchase housing in relatively affluent council estates, it has also been observed that the process has led to a 'ghettoisation' effect where many of the better council estates are now almost entirely owner-occupied leaving councils with only the less desirable housing stock in more problematic estates. From a modelling viewpoint, the notion of 'clustering' in housing sales is of interest here. The observations made above would suggest that there should be at least some autocorrelation in rates of homeownership—it is often the case that council estates are perceived as a whole, so that exercising the right to buy council housing will be manifested in spatial clusters.

However, these observations apply only to *council* housing and there is also a sizeable private rented sector. In some areas the ghettoisation model may still apply, but there are other areas, particularly those with a high proportion of wealthier short-term residents, where affluent privately rented homes exist alongside owner-occupied homes, and where there is less of a tendency to cluster. It may also be the case that in some geographical regions there is not as strong an 'owner-occupier' culture as in others so that even in the more desirable estates there is still a mix of rented and owned housing, again exhibiting less of a tendency to cluster.

The above serves to suggest that levels of homeownership may exhibit different levels of clustering (or autocorrelation) in different areas. Using a technique such as that described above provides a means of exploring this phenomenon by fitting locally varying estimates of ρ over a range of geographical space. Linking maps of this variation to other socioeconomic indicators in the region should provide some insight into the process of spatial variations in autocorrelation.

The data are the 1991 proportions of owner-occupied households as a percentage of all households for census wards in Tyne and Wear (OPCS, 1992). These values are mapped in figure 1. A value of 0.5 has been subtracted from each proportion so that the data have a mean close to zero which facilitates the fitting of equation (1). Hence, the index is positive if the majority of housing in a ward is owner-occupied and negative if it is not. The pattern is one where rented accommodation dominates the wards towards the city centre and along the river Tyne and owner-occupied housing dominates in the peripheral wards. The autocorrelation in model (1) can be interpreted as assuming that areas surrounded by higher rates of owner-occupation are likely to have higher rates themselves, and vice versa for high rates of rented accommodation. Applying the above
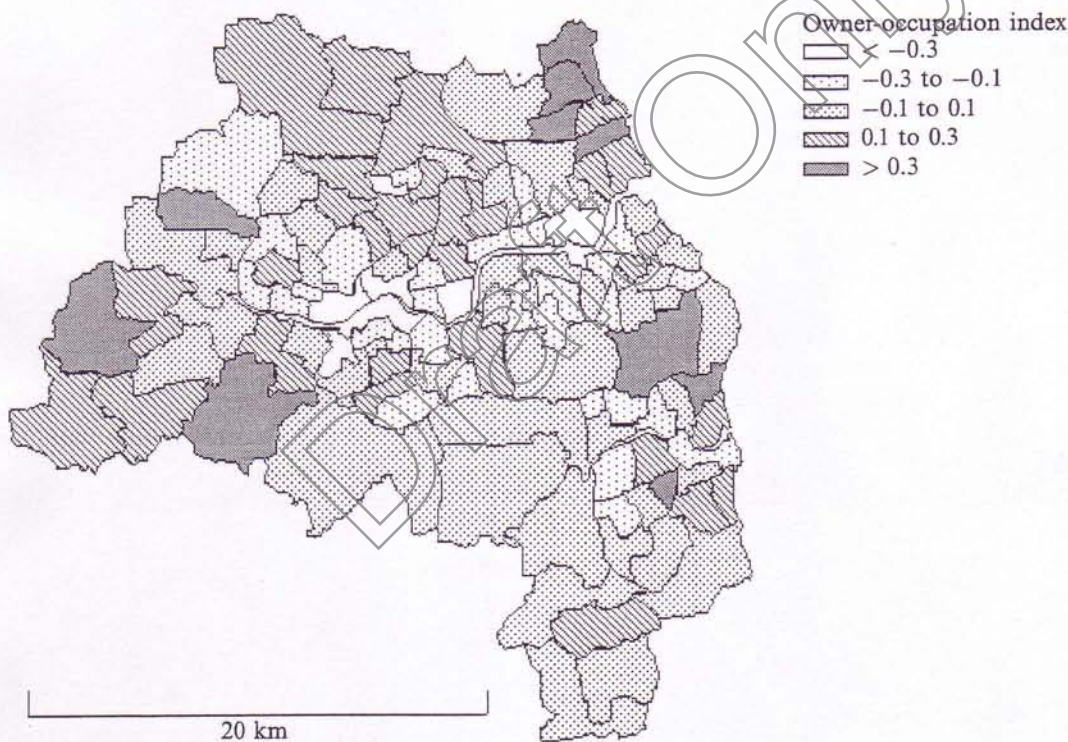

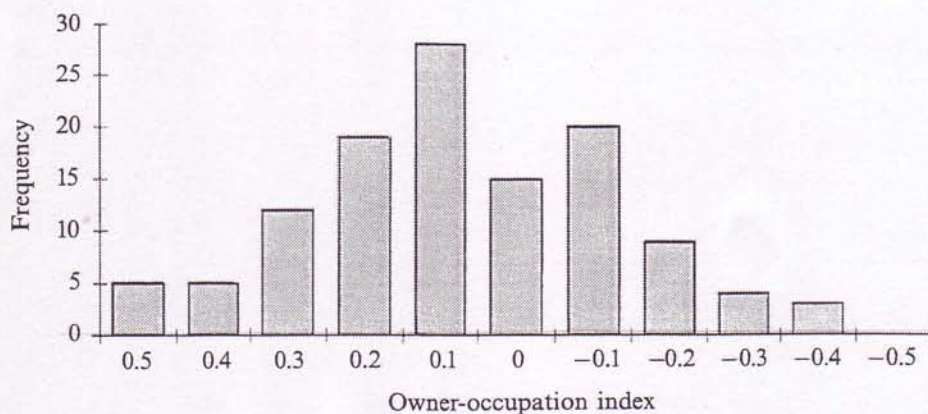**Figure 1.** Levels of home owner-occupation in Tyne and Wear.

Figure 2. Histogram of the owner-occupation index displayed in figure 1.

transform gives the histogram shown in figure 2. From this it may be seen that zero is a plausible central point here and that the index follows a roughly normal distribution. From this standpoint it is reasonable to fit either of models (1) or (2).

Fitting model (1), a global Ord model, produces an estimate of $\rho = 0.55$ and a variance of 0.031. A graph of predicted versus actual owner-occupation rates is given in figure 3 and a map of the residuals in figure 4. From this latter map it may be seen that prediction errors seem to have greater magnitude in some regions than in others—suggesting perhaps that a smoothed index of owner-occupation is a better predictor of actual owner-occupation in some areas than in others.
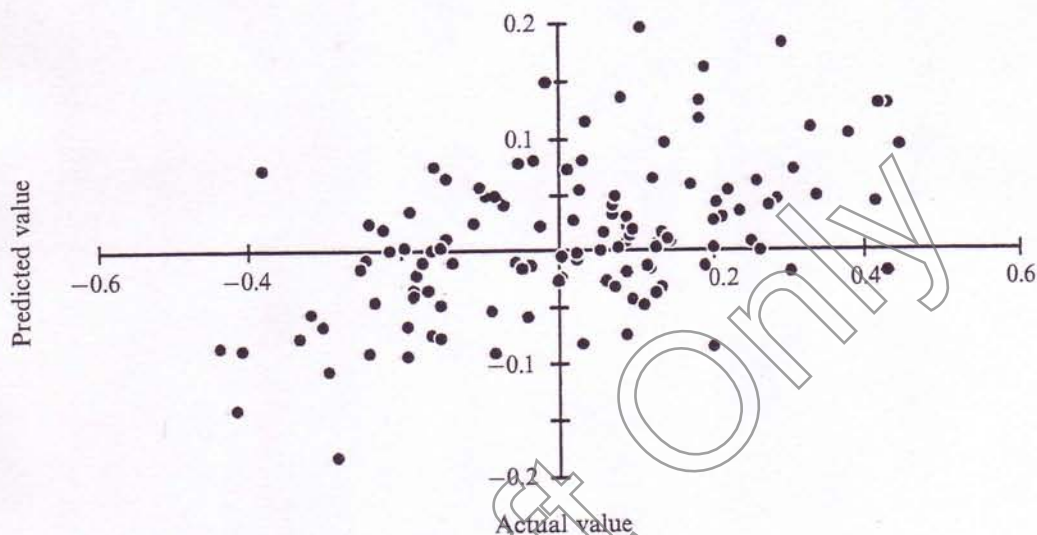

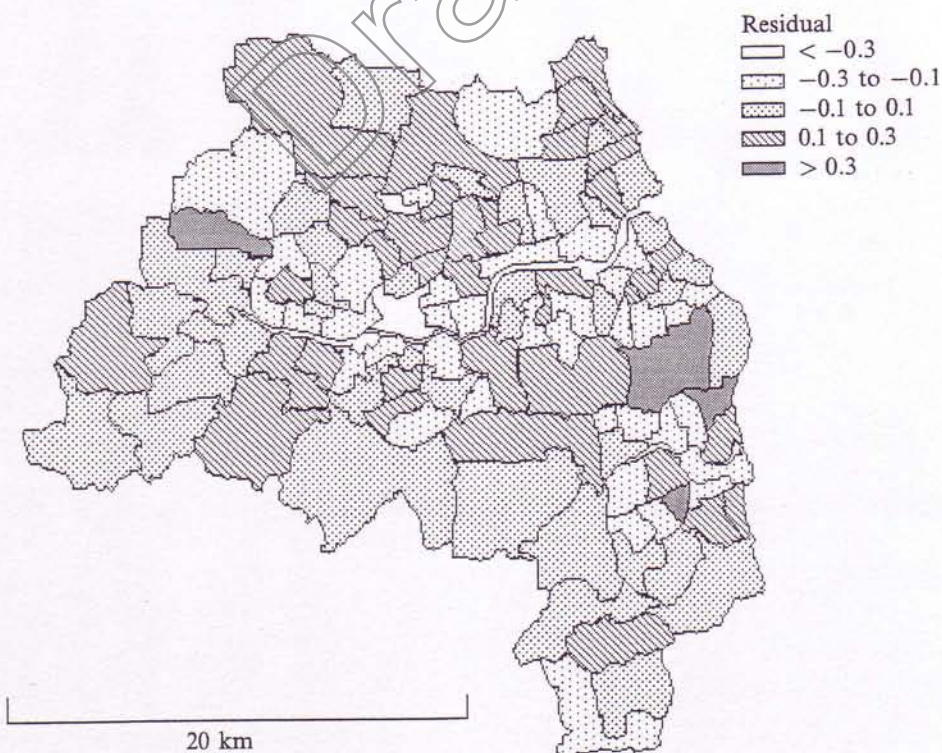**Figure 3.** Predicted versus actual owner-occupation levels.

Figure 4. Ord model residuals.

The next stage is to consider a geographically weighted model for ρ and for this a kernel function must be chosen. Initially, a Gaussian kernel is chosen:

$$K(d) = \exp\left(-\frac{d^2}{k^2}\right),\tag{20}$$

where d is the distance between the centroid of each ward and the point at which ρ is to be estimated. Initially k was chosen to be 5 km so that the weighting of zones 5 km from the point of estimation will be about one third (compared with unity in an unweighted estimation) and zones more than about 10 km away from i will have negligible influence on the estimation of ρ_i. Estimating ρ_i at the centroids of each census ward (and shading in the ward according to this value) gives the map in figure 5. Clearly, some degree of variation in the estimated values of ρ_i is evident with some zones having considerably higher values than others. Plotting ρ_i against
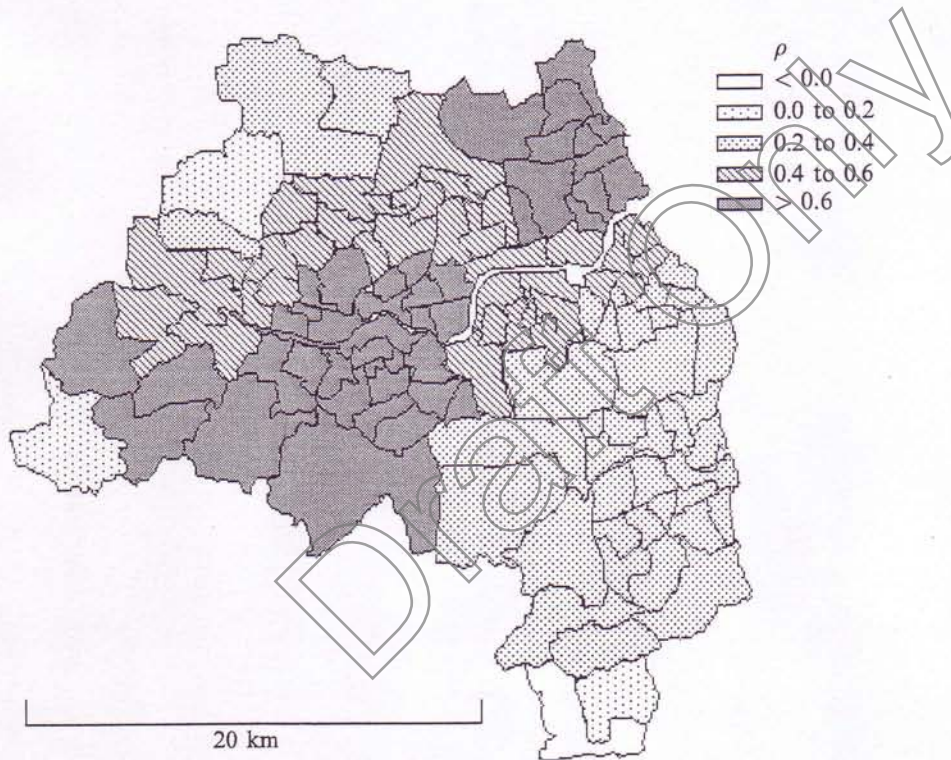

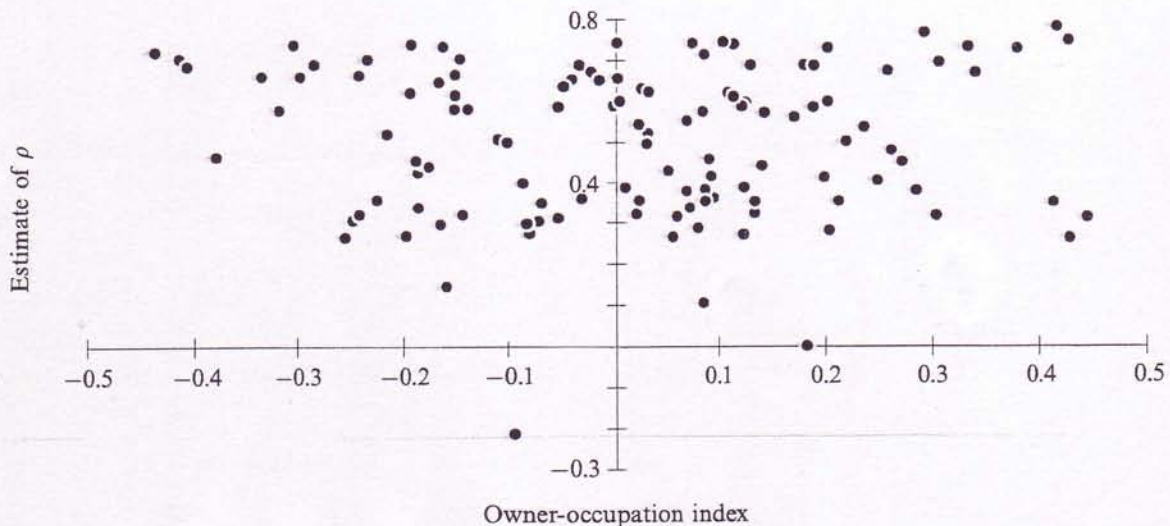**Figure 5.** Geographically weighted ρ (Ord model).

Figure 6. Graph of owner-occupation against local estimate of ρ.

levels of owner-occupation (figure 6) shows two groups of wards: those, primarily the inner-city wards, where low owner-occupation rates are combined with a high spatial autocorrelation; and those, primarily in the northeastern coast area and southwestern areas where high owner-occupation rates are associated with equally high spatial autocorrelation. Other areas of the region show less evidence of geographical cohesion.
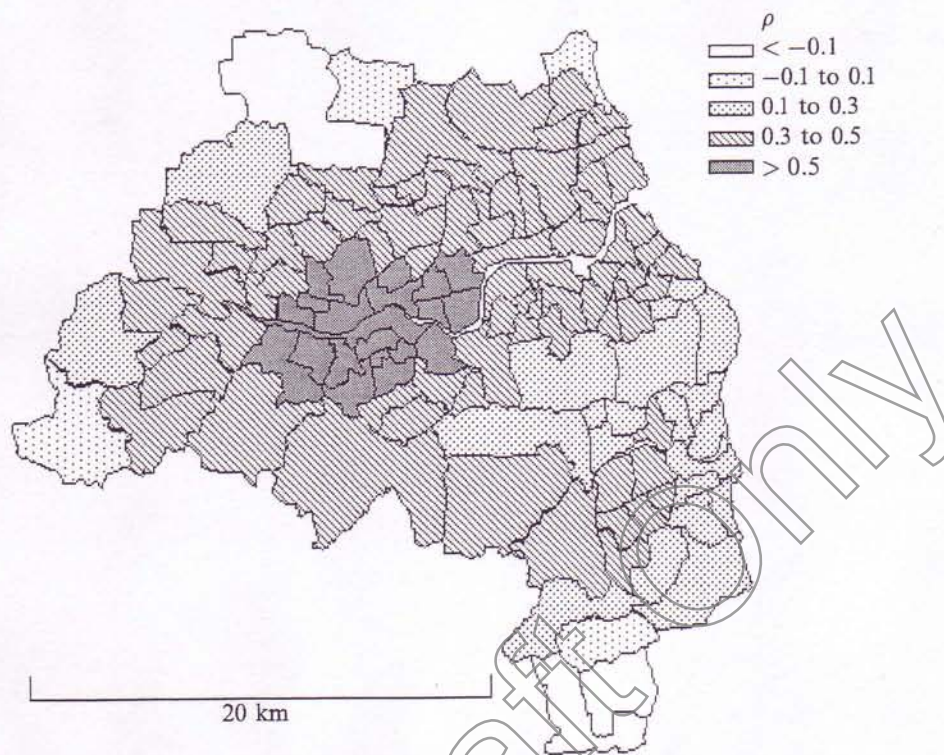

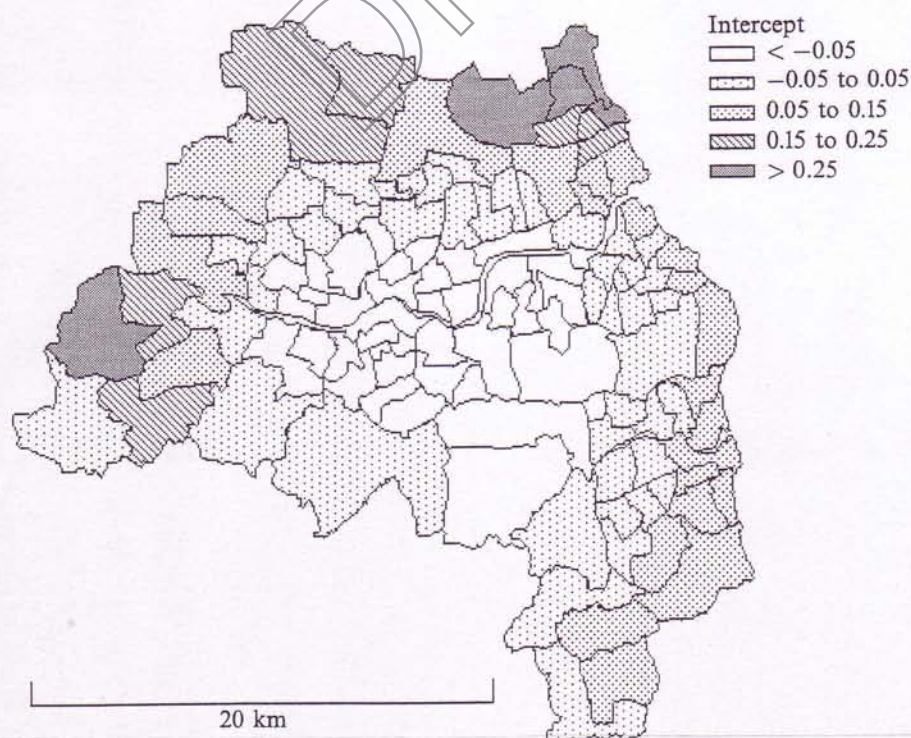**Figure 7.** Estimates of ρ for model with intercept term.

Figure 8. Intercept map for model with intercept term.

A final analysis of these data is carried out with model (6), the model in which the mean is calibrated rather than being input exogenously. Using the same kernel as before, we can estimate localised estimates of both ρ_i and β_i as in equation (19) and these are shown in figures 7 and 8. The map for ρ_i shows similar patterns to that for the simpler geographically weighted model. The range of values for β_i and their magnitudes suggests that, once the proportions of homeownership have been centred around 50%, the zero mean assumption is not unreasonable.

Further issues 1: choosing a kernel bandwidth

Although the above example demonstrates how techniques proposed in this paper may be used to explore the spatial variation in spatial association and investigate the extent to which spatial diffusion processes vary between areas, there are still several issues which need to be addressed. Two of the most pressing problems are the choice of kernel bandwidth and the provision of a formal test as to whether global models hold. In this section attempts to address both of these questions will be made.

First consider the problem of kernel bandwidth. Clearly, the choice of k in equation (20) will affect the overall pattern of local calibration. For example, if model (2) is recalibrated with a k-value of 10 km (see figure 9), it can be seen that the spatial distribution of ρ_i appears smoother than the original in figure 7. Although both maps display similar patterns, it would be helpful if some optimal value of k could be found. Often in exploratory model fitting a simple paradigm such as a least-squares fit may be applied. In the case of model (2) this is equivalent to minimising the squared differences of the observed y-values and the smoothed y-values multiplied by a locally varying estimate of ρ with respect to k. In each case the model is calibrated on the centroid of the ward for its respective y-value. An alternative but similar approach is to find k maximising the likelihood of each observed y. There is, however a problem, similar to that found by Brunsdon et al (1996), with these approaches. Suppose k is allowed to become extremely small so that the influence of all y-values *except* the one in question is neglible. In this case a perfect fit can be obtained by setting ρ_i to be the
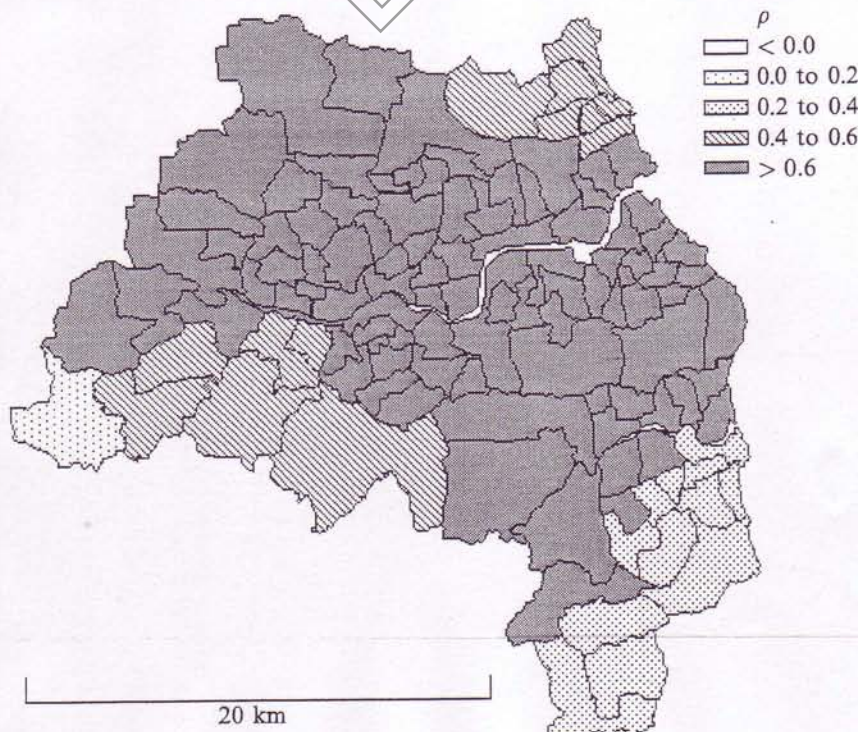

Figure 9. Estimates of ρ made by using a larger bandwidth.

ratio between the actual and smoothed y-values for that zone. Thus, in terms of least squares (and maximum likelihood), the 'best' fit corresponds to a zero bandwidth which is clearly unsatisfactory. Applying similar techniques to those of Brunsdon et al (1996), one way to address this problem is to use a cross-validation scoring technique. In this method, each y-value is omitted from the data set, which is then calibrated, and then a fitted y-value based on this calibration is used for a least-squares goodness-of-fit statistic. This is not unreasonable because it is perfectly possible to estimate a ρ_i value at any point in the study area. Thus, if a given zone is omitted from the calibration process, it is still possible to estimate ρ_i at its centroid and then to compute a fitted value of y_i.

One major alteration that will need to be made, however, is to the **W** matrix. This contains information about the connectivity (or some other facet of spatial arrangement) for *all* of the zones, including the one omitted. For the 'leave-one-out' model, **W** must be modified by striking out the ith row and column if zone i is to be omitted. After this, the rows must be restandardised to sum to one if **W** is to remain a mean-smoothing matrix. Although all of this is possible, it does require a new set of eigenvectors to be computed for each calibration of the model with a concomitant increase in computing overheads. As the sum of squared errors will require computation several times in order to find an optimal k, it is worth computing all n lists of eigenvalues (one for each zone omitted) in advance and choosing the appropriate one in each computational case. An alternative is not to consider the cross-validation error for *all* zones but to use a subset of zones. This could either be strategically chosen (with some from urban areas and some from rural areas) or taken as a random subsample of zones.

With the random subsample method applied to the owner-occupation data, figure 10 shows the relationship between the sum of squared errors and k for the 10 wards shown in figure 11. This would suggest that, with the cross-validation approach to choice of k, an optimal value is about 5.5 km. Again, it should be noted that this does not mean that zones 5.5 km away from a point of estimation are ignored but that they are downweighted to about one third of full weighting. Zones up to about 11 km away from point i contribute to the local calibration of ρ_i in some nonnegligible way.

Another approach to choosing suitable k-values by means of cross-validation, avoiding the computation of several sets of eigenvalues, would be to calibrate the model (2) or model (6) with least squares techniques instead of maximum likelihood. This is equivalent to calibration with an ordinary regression model using a smooth y-variate as though it were a predictor variable and ignoring the effects of spatial autocorrelation. This can be calibrated with the geographically weighted regression (GWR) techniques of Fotheringham et al (1997b; 1997c) and Brunsdon et al (1996)—in this case, ρ_i is the
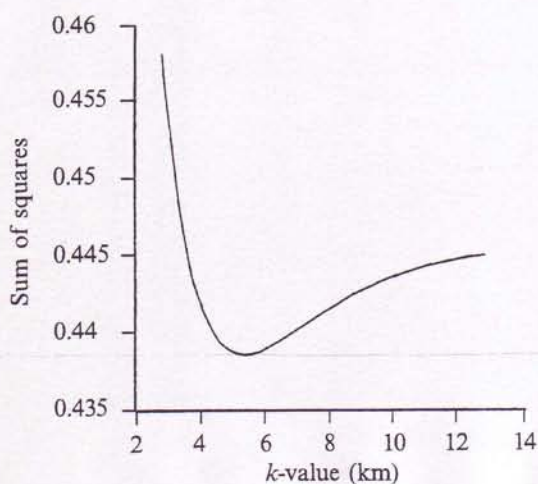

Figure 10. Cross-validation scores.

Figure 11. Sample points for significance test.

equivalent of the coefficient of the smoothed y-value. Finally, with the cross-validation techniques introduced for GWR, it is possible to select an optimal bandwidth.

The advantage of the above method is that it is effectively optimising the second term in (14)—the one without the eigenvalues. However, this is not without a price. The inconsistency of the least squares estimator of ρ has been noted by Whittle (1954) and it may also be shown that its asymptotic efficiency (the ratio between the variance of the maximum likelihood estimator and the least-squares estimator) is also poor for larger values of ρ_i. The main hope here is that, although the actual values of ρ_i provided by this method may be poor, the estimate for an optimal k may still be reasonable. This area will require further investigation.

Further issues 2: testing stationarity of the autoregressive coefficient
Another issue which needs to be addressed is that relating to formal tests of non-stationarity. Clearly, a nonstationary autoregressive model is more complex than a stationary one and if there is only poor evidence of nonstationarity then effort would be expended attempting to interpret essentially random fluctuations in ρ_i. Any analysis of nonstationarity should be preceded by some formal significance test. This test could be based on the null hypothesis that

$$\rho_i = \text{constant for all } i, \tag{21}$$

that is, the only fluctuations observed in ρ_i are those attributable to chance. This is a relatively difficult test to implement because the sampling distribution for the estimates of ρ_i is not known analytically. As the estimates used are maximum likelihood, the asymptotic distribution of ρ_i is well known so tests comparing ρ_i values in two different subregions of study may be a valid approach. However, it is not determined for a given arrangement of zones whether asymptotic conditions have been approximately achieved. To paraphrase Besag (1974), it is difficult to imagine the number of census wards in Tyne and Wear tending to infinity, but a further problem is that it is equally difficult to tell whether the actual number of wards in Tyne and Wear is sufficiently

close to infinity to apply tests based on asymptotic assumptions. Further objections to this technique could be raised on the grounds that, if geographically weighted estimates of ρ_i had already been mapped, this would be a post-hoc rationalisation of the null hypothesis. A final objection is that finding that the chosen subregions do not differ significantly in ρ_i does not imply that ρ is fixed everywhere. It is possible that a different choice of subregions may have successfully identified differences. This suggests that such a test is not very powerful in testing the full implications of the logical negation of hypothesis (21).

Monte-Carlo or Hope testing (Hope, 1968) might be a more fruitful approach because it does not rely on the parametric assumptions for the sampling distribution of ρ_i required by the method suggested above. To counter the other two objections, ρ_i should be measured at a series of sample points spread homogeneously across the study area and the variability of these estimates used as a test statistic for hypothesis (21). This approach, by covering most of the study area, should be better able to detect deviation at an unknown point in the study area. Also, because a reasonably uniform distribution of sampling points is used, one cannot be accused of comparing a post-hoc choice of subregions.

There are some issues that must be dealt with here. First, as with the approach suggested for choosing the k-value, it is important to choose the sampling points wisely. In particular, a set of sampling points representing both urban and rural areas should be used. Two approaches are suggested: one is that sampling could be based on centroids of wards (or whatever areal units are used in a given example); the second is that it could be based on a regular lattice covering the study area. In either case, care should be taken to ensure that the sample of points is in some way representative of the geography of the area under investigation. One possibility when using wards may be to select sample centroids randomly by means of a sampling process in which the probability of selection for a given zone is proportional to the population of that zone.

A further matter arising relates to the Monte-Carlo methodology itself. In previous works the authors have used a randomisation-based methodology for tests of spatial association (Brunsdon et al, 1996; Fotheringham et al, 1997c). In this approach, attributes of zones are permuted randomly amongst the zones in each randomisation. If a statistic, such as a Pearson correlation coefficient, is space invariant, then such randomisation brings no change. If a statistic is dependent on the spatial arrangement of the data—such as Moran's I-statistic (Moran, 1950)—then clearly some variation will be observed. However, under a null hypothesis of no spatial association in the data, any of the permutations obtained would be equally likely so that the variation in the statistic will be a result of its null distribution. Comparing the *actual* value of the statistic with a sample of permutation-based values forms the basis of the Hope test.

There are, however, difficulties with this technique in the current case. The main problem is that the null hypothesis here—equation (21)—is not one of no spatial association, but of a *uniform degree of spatial association*. What this suggests is that the 'random permutation' approach is no longer valid. The null hypothesis in this case demands that one form of spatial association (the uniform model) is tested against another (the nonuniform model). It is an irrelevance to compare the degree of variation of ρ_i in the observed sample against the degree of variation expected were there no spatial pattern at all. How, then, might a more desirable test be devised? One possibility is to identify a phenomenon which would have no spatial association under hypothesis (21), and permute this, rather than the observed zonal attributes. This is indeed a possibility because equation (7) shows that e is a matrix transformation of y—but of course e is a set of independent variables. That is, under a null hypothesis that equation (1) holds,

it may be stated that

$$y = \mathbf{A}^{-1}e,$$

or

$$y = (\mathbf{I} - \rho\mathbf{W})^{-1}e. \tag{22}$$

Thus, by permuting e it is then possible to derive a vector of variables y which may then be used for the Monte-Carlo simulation. The variance of the estimated ρ_i values in these simulations will then provide a reference distribution with which to compare the observed variance of ρ.

Thus, the Monte-Carlo algorithm for testing hypothesis (21) may be set out as below:

1. Estimate ρ_i at the sample points.
2. Compute the variance of the estimated values obtained above and call it v.
3. Compute $e = \mathbf{A}y$.
4. Compute e^*, a random permutation of e.
5. Compute $y^* = \mathbf{A}^{-1}e$.
6. Estimate ρ_i at the sample points, based on y^*.
7. Compute the variance of the estimated values obtained above.
8. Repeat steps $4-7$ $n-1$ times.
9. Compare v against the $n-1$ randomised values to yield a significance level.

A further refinement could be achieved by employing Besag and Clifford's (1991) method for sequential Monte-Carlo testing. In this approach the test is applied until either l tests give a test statistic more strongly opposed to the null hypothesis than the actual statistic, or a given number of simulations have been achieved. Typically l would be 10 or 20. In cases where the null hypothesis is true this tends to reduce the number of simulations needed to perform the test, and so improves computational efficiency.

A final observation must be made about the use of ρ_i in this algorithm. Strictly, the null hypothesis in this test is that equation (1) holds *for a known value of* ρ. That is, it tests, say $\rho = 0.5$ everywhere against a hypothesis that $\rho \neq 0.5$. In practice, the value of ρ will not be known in advance but an estimate will be computed with methods given in this paper. This suggests that there will be some extra level of variability in the sampling distribution of the variance of the local estimates of ρ, because of the standard error of the global ρ estimate. Thus, a test which is *just* significant should be treated with some caution. It is hoped that in future studies, a more satisfactory approach to this testing procedure may be developed.

The results of applying this test to the owner-occupation data are shown in table 1. From this table it may be seen that the observed variance, at 0.229, is in the lower tail of the randomisation distribution, ranking 98th in 100 trials. This suggests that there is some justification in the adoption of a spatially nonstationary Ord model.

Table 1. Monte-Carlo simulation results to test for nonstationarity: variance of ρ_i.

Distribution	Variance
Observed	0.229
Experimental	
minimum	0.060
lower quartile	0.123
median	0.155
upper quartile	0.191
maximum	0.234

Further issues 3: adding explanatory variables

The Ord model, as used above, is a useful tool for modelling the degree to which a given variable exhibits spatial association. However, although spatial self-association plays an important rôle in many geographical processes, it is seldom this alone which explains variations in many phenomena. For example, although rates of owner-occupation in surrounding areas may well have some bearing on the rate of owner-occupation in a given area, there may well be other factors such as unemployment levels or household composition which would also affect levels of owner-occupation. For this reason, a logical extension of the methodology is to incorporate explanatory variables in the model, as in equation (15). A simple example of this is considered above when an offset is added to the model.

Applying a geographically weighted approach to this model gives a regression model in which coefficients vary over space and the degree of autocorrelation alters. This would suggest, in the owner-occupation versus unemployment model discussed above, that not only could the effect of the surrounding level of owner-occupation vary over space but also that the linkage between the rate of owner-occupation and unemployment may exhibit spatial nonstationarity. The latter may be caused by the actual price of housing—if owning housing is relatively cheap in one region then there may be a greater take-up rate than in other more expensive areas having similar levels of prosperity.

An analysis of the model suggested above is carried out in figures 12 and 13—showing the ρ_i estimates and the coefficient for male unemployment in the model, respectively. It can be seen that the levels of autocorrelation vary in a slightly different way once levels of male unemployment have been accounted for and that the coefficient for male unemployment is steepest in the central areas. As with the simple Ord model, it is possible to use a cross-validation method to choose an optimal k-value and to test hypotheses of spatial nonstationarity.
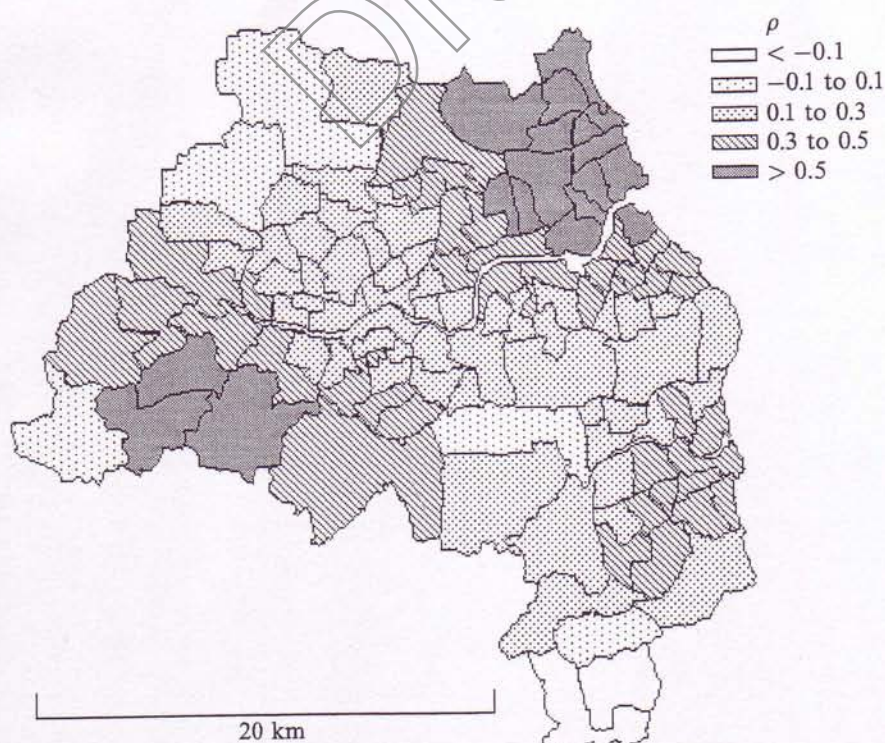

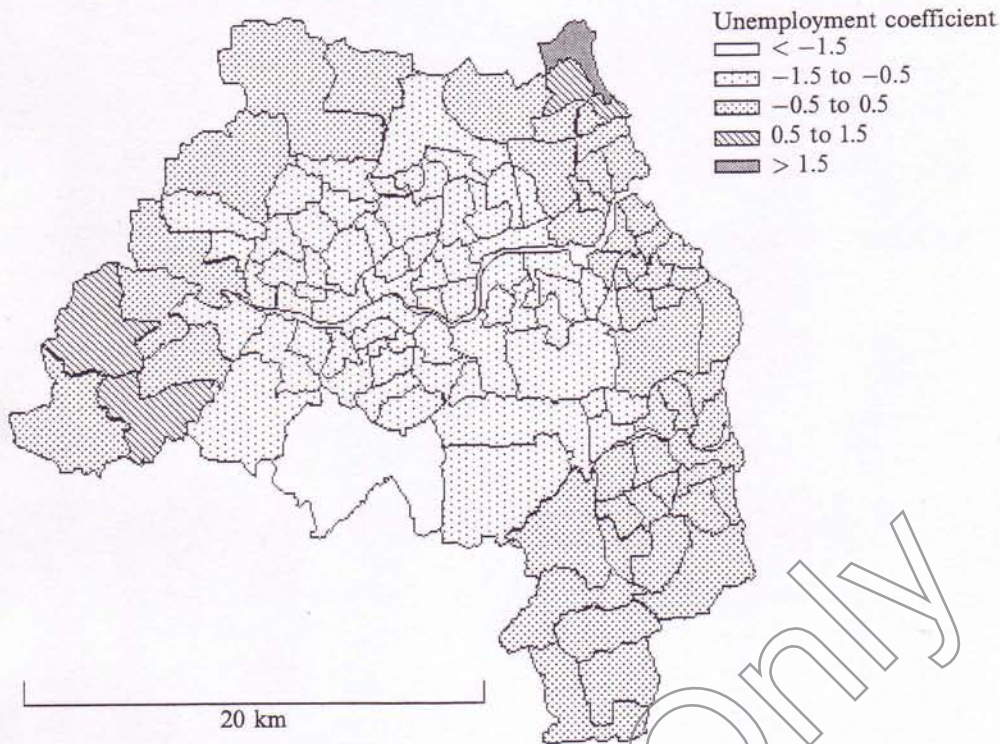**Figure 12.** Estimate of ρ mode from model 3.

Figure 13. Unemployment coefficient from model 3.

Conclusions

In this paper we explore the possibility of examining spatial nonstationarity in a particular form of spatial regression, the Ord model, although the technique can easily be applied to any other framework in which estimates of spatial autocorrelation are obtained. Our findings include a method for calibrating such a model and some tentative approaches to testing the model against a stationary null hypothesis. It should be noted that at present the development of such tests is in its early stages and it is hoped that more robust methods may suggest themselves after a more rigorous analytical treatment of the problem. For example, one notable characteristic of the calibration of the Ord model with spatially reweighted maximum likelihood is the high computational overhead. This could be reduced if a least squares approach were adopted. In this study the maximum likelihood approach was chosen on the grounds that this would provide a consistent estimate in the case where a global Ord model did hold (assuming the kernel bandwidth remained fixed as the sample size increased); a least squares estimator would not ensure this consistency. However, it is hoped that more detailed study of the behaviour of the two kinds of estimator should eventually lead to a less ad hoc choice of estimator.

To gain a greater understanding of the spatially varying autoregressive model described here, a comparison with the work of Anselin (1995) might be helpful. Whereas Anselin disaggregates the weighting matrix into spatial components (a local area decomposition) and examines the influence of these local weighting matrices on the global model, in our method each zone may be thought of as responding to a different model so that the autocorrelation measure varies over space. The essence of our model, therefore, is not to search for any global statement of relationships but to examine local variations in such relationships.

We feel that this approach represents an important breakthrough. The advent of GIS has brought about a much greater awareness that spatial location is often a very important explanatory factor in geographical processes (a fact quite amazingly ignored

in many analyses of geographical data), and this had led more people to turn to *spatial* methods of data analysis and given rise to publications such as that by Fotheringham and Rogerson (1993). There is currently a great interest in 'local' as opposed to 'global' statistics and spatial variations are increasingly being recognised as important facets of analysis with spatial data, rather than as irrelevances to be ignored in the search for general laws. This paper adds to the growing literature in the field of truly spatial statistics.

References

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

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

Besag J E, 1974, "Spatial interaction and the statistical analysis of lattice systems" *Journal of the Royal Statistical Society* B **36** 192 – 236

Besag J E, Clifford P, 1991, "Sequential Monte-Carlo *p*-values" *Biometrika* **78** 301 – 304

Brundson C F, Fotheringham A S, Charlton M, 1996, "Geographically weighted regression— a method for exploring spatial non-stationarity" *Geographical Analysis* **28** 281 – 298

Casetti E, 1972, "Generating models by the expansion method: applications to geographic research" *Geographical Analysis* **4** 81 – 91

Fotheringham A S, 1997, "Trends in quantitative geography I: stressing the local" *Progress in Human Geography* in the press

Fotheringham A S, Pitts T C, 1995, "Directional variation in distance-decay" *Environment and Planning A* **27** 715 – 729

Fotheringham A S, Rogerson P, 1993 *Spatial Analysis and GIS* (Taylor and Francis, London)

Fotheringham A S, Brunsdon C, Charlton M E, 1997a, "Geographically weighted regression: a natural evolution of the expansion method" *Environment and Planning A* in the press

Fotheringham A S, Brunsdon C, Charlton M E, 1997b, "Two techniques for exploring nonstationarity in geographical data" *Geographical Systems* **4** 59 – 82

Fotheringham A S, Brunsdon C, Charlton M E, 1997c, "Measuring spatial variations in relationships with geographically weighted regression", in *Recent Developments in Spatial Analysis* Eds M M Fischer, A Getis (Springer, New York) chapter 4, in the press

Getis A, Ord J K, 1992, "The analysis of spatial association by use of distance statistics" *Geographical Analysis* **24** 189 – 206

Greig D M, 1980 *Optimisation* (Longman, Harlow, Essex)

Hope A C A, 1968, "A simplified Monte Carlo significance test procedure" *Journal of the Royal Statistical Society* B **30** 582 – 598

Jones J-P III, Casetti E, 1992 *Applications of the Expansion Method* (Routledge, London)

Mead R, 1967, "A mathematical method for the estimation of interplant competition" *Biometrics* **23** 189 – 205

Moran P A P, 1950, "Notes on continuous stochastic phenomena" *Biometrika* **37** 17 – 23

OPCS, 1992 *1991 Census: Definitions: Great Britain* Office of Population Censuses and Surveys (The Stationery Office, London)

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" *Geographical Analysis* **27** 286 – 306

Whittle P, 1954, "On stationary processes in the plane" *Biometrika* **41** 434 – 449