Chapter 9
Spatial Modelling

9.1 Introduction

Spatial model is (usually) a mathematical description of the structure of spatial phenomenon that makes it possible to simulate it (usually) in a computer environment, to measure the effect of external factors.

In urban analysis, spatial modelling is quite important because it enables us to compare plans by simulating their effect on urban structure.

Spatial models are built on the knowledge about spatial phenomena obtained by spatial operation and analysis. Spatial modelling, therefore, is usually done after spatial analysis.

References - General spatial modelling

- Griffith,
- Fotheringham

Spatial models are classified by

1. whether they are behavioral or descriptive,
2. whether they are deterministic or probabilistic.

Behavioral model is a model built on a set of facts and reasonable assumptions. It is necessarily theoretical. It often uses the data obtained in experiments and observations; some models are completely deductive since they do not use observed data.

Descriptive model is an inductive model built on the data obtained in experiments and observations. Its objective is to describe the data as accurate as possible, and to understand their underlying structure. Descriptive model is often used for exploratory purposes.
9. Spatial Modelling

Deterministic and probabilistic models

Deterministic model is a model in which spatial phenomena are described deterministically, that is, without any uncertainty. It is based on the assumption that a set of theories can describe spatial phenomena perfectly.

Probabilistic model is a model in which spatial phenomena are described as stochastic, uncertain, and indeterministic. It describes spatial phenomena by a combination of theories and probabilistic errors that represent the effect of unknown factors.

General procedure of model building

1. Consider a probabilistic model based on a set of facts and reasonable assumptions.
2. Obtain the data from experiments and observations.
3. Estimate the parameters of the model.
4. Test the significance of parameters and the model.
5. If several models are considered, compare them and choose the one of the best fitness.

Examples of spatial models

1. Regression-based models
2. Spatial choice models
3. Point processes
4. Spatial diffusion models

Note

Spatial models have been developed in various fields related to spatial analysis, say, geography, epidemiology, economics, ecology, seismology, archaeology, transportation science, and so forth.

Because of this, it is difficult to explain spatial models in a well-structured, systematic way, though I will try to explain them systematically.

9.2 Regression-based models

The regression-based model is a descriptive model of spatial phenomena. It describes an attribute of spatial objects by a set of other attributes, where spatial autocorrelation exists among attribute values.

Example:
Modelling of population density by land use ratios, distance from city centers, land prices, etc.
9. Spatial Modelling

### References - Regression models


9. Spatial Modelling

Regression-based models can be applied to points, lines, and polygons, but they are mainly used for tessellations, a set of polygons that exclusively covers a region.

They explain an attribute of a tessellation by a set of other attributes.

For the present, we treat only continuous numerical attributes whose domain is \((-\infty, \infty)\).

9.2.1 Ordinary regression model

The ordinary regression model is represented by

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \quad (i = 1, \ldots, n)
\]

where \(\epsilon_i\) is the probabilistic error term.

This set of \(n\) equations is represented by a single equation.

Variables represented by vectors and matrices:

\[
Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}
\]
Ordinary regression model assumes
\[ Y = X\beta + \varepsilon \]
and estimate the parameter vector \( \beta \).

There are numerous variations of regression models. They are different chiefly in the treatment of the probabilistic error term \( \varepsilon \).

In the ordinary regression model, the error term \( \varepsilon \) is assumed to follow the identical (not necessarily normal) probability distribution of mean 0 and variance \( \sigma^2 \) independently.

Mathematically, it is written as
\[
E[\varepsilon] = 0 \\
V[\varepsilon] = E[\varepsilon \varepsilon^T] = \sigma^2 I
\]

The second term \( V[\varepsilon] \), the matrix of variances, is called the variance-covariance matrix, and is denoted by \( C \).

Ordinary Least Square (OLS) method

Under this assumption the parameter vector \( \beta \) (and consequently the whole model) is estimated from observed data by the ordinary least square method. This method calculates the parameter vector \( \hat{\beta} \) that minimizes the square sum of the estimation error:
\[
\min_{\beta} \sum_{i=1}^{n} e_i^2 \iff \min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i0} - \cdots - \beta_k x_{ik})^2
\]

This problem is rewritten as
\[
\min_{\beta} e' e \iff \min_{\beta} (Y - X\hat{\beta})' (Y - X\hat{\beta})
\]

This estimator is optimal in the sense that it is unbiased and gives the least estimation variance.
The parameter vector $\hat{\beta}$ is given by solving a set of equations

$$\frac{\partial}{\partial \hat{\beta}} (Y - X\hat{\beta})' (Y - X\hat{\beta}) = 0$$

The estimator of $\beta$ is

$$\hat{\beta} = (X'X)^{-1} X'Y$$

The estimator of the variance of error term is given by

$$\hat{\sigma}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - (k + 1)}$$

Note that the expectation of error term is zero.

**Maximum Likelihood (ML) method**

Another approach for parameter estimation is the maximum likelihood method. To use this method, we assume that the error term follows a normal distribution $N(0, \sigma^2)$:

$$\epsilon_i \sim N(0, \sigma^2)$$

Therefore,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_i - \hat{y}_i)^2}{2\sigma^2} \right)$$

The likelihood is given by

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(y_i - \hat{y}_i)^2}{2\sigma^2} \right)$$

To calculate the parameters that maximize the likelihood, we usually convert it to the log likelihood and solve

$$\max_{\beta, \sigma^2} \log L$$

$$\max_{\beta, \sigma^2} \left[ \frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (Y - X\hat{\beta})' (Y - X\hat{\beta}) \right]$$

Therefore, in this case, the maximum likelihood method is equivalent to the ordinary least square method. The estimator of $\hat{\beta}$ is thus maximum likelihood estimator.
Model building and estimation can be done as shown above. If the estimators and the model are to be tested, an additional normality assumption is necessary on the error term $\varepsilon$. When $\varepsilon$ follows a normal distribution, the $t$ statistic is calculated for each parameter element in $\beta$ and its significance can be tested by the $t$-test.

### 9.2.2 WRM: Weighted Regression Model

The ordinary regression model makes a strong assumption that the error term $\varepsilon$ follows the identical probability distribution of mean 0 and variance $\sigma^2$ independently.

The weighted regression model relaxes this assumption, that is, it permits that the error terms have different variances.

The weighted regression model is written as

$$ Y = X\beta + \varepsilon $$

which is the same as that of the ordinary least square method. The difference lies in the variance-covariance matrix $C$ of the error term $\varepsilon$.

This indicates that the error terms independently follow probabilistic distributions of different variances.

To estimate this model, we use the weighted least square method. The ordinary least square method treats all the estimation errors equally. Therefore, estimation error for an object of a large variance is more influential on the estimation result than that of a small variance.

The weighted least square method, on the other hand, estimates the unknown parameters using the square sum of weighted estimation error.

$$ w_i \cdot \text{Weight for object } i $$

The weighted least square method solves

$$ \min \sum w_i y_i^2 \iff \min \sum w_i (y_i - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k)^2 $$
9. Spatial Modelling

Using the vector-matrix representation

\[
W = \begin{bmatrix}
w_1 & 0 \\
0 & w_n
\end{bmatrix}
\]

we have

\[
\min \sum w_i e_i^2 \Leftrightarrow \min \{Y - X\hat{\beta}\} \cdot W \{Y - X\hat{\beta}\}
\]

The parameter vector \(\hat{\beta}\) is given by solving a set of equations

\[
\frac{\partial}{\partial \beta} (Y - X\hat{\beta})^T W (Y - X\hat{\beta}) = 0
\]

The estimator of \(\beta\) is

\[
\hat{\beta} = (X^T W X)^{-1} X^T W Y
\]

9. Spatial Modelling

In the weighted regression model, we have to determine the weights. If \(\sigma_i^2\), the variance of dependent variable \(y_i\) is known, we usually determine the weight as

\[
w_i = \frac{1}{\sigma_i^2}
\]

and thus

\[
\hat{\beta} = (X^T C X)^{-1} X^T C Y
\]

In general, however, the variances or their estimators are not known, so we have to determine the weights by other methods.

If the dependent variable is an average of several observations, we may use the number of observations as the weights.

If often happens that the variance increases with independent variables. If so, we may use the reciprocal of independent variables as the weights.

9.2.3 GRM: Generalized Regression Model

These methods may seem arbitrary. Moreover, there are the cases where no information is available about the variances.

In such a case, in spatial analysis, we use the covariogram function to the estimation of \(C\), and estimate \(\beta\) and \(C\) by an iterative process. This method will be discussed in the next section.

The generalized regression model further relaxes the assumption of the weighted regression model: It permits that each error term \(\varepsilon\) follows different probability distributions, and that there exist a correlation among the error terms.
9. Spatial Modelling

Note that not only diagonal elements of C but also other elements are not equal to zero. This means that a correlation exists among error terms, and its strength is indicated by the values of elements in C.

The generalized regression model is estimated by the generalized least square method. The procedure, however, is quite similar to the weighted least square method except the off-diagonal elements of C are nonzero. Therefore, the problem and the estimator of $\beta$ are

$$\hat{\beta} = (X'C^{-1}X)^{-1}X'C^{-1}Y$$

This, however, is not a complete solution to the problem of model estimation, since the result includes unknown variance-covariance matrix C. We have to estimate both $\beta$ and C simultaneously in order to estimate the full model. To this end, we use an iterative approach which uses the covariance function of the dependent variable between spatial objects.

Estimation procedure

1. Assume that the error terms follow an identical probability distribution of mean 0 and variance $\sigma^2$ independently, that is, $C = \sigma^2 I_n$. Then apply the ordinary regression model to the data and calculate the estimator of observation Y.

2. Calculate the residuals of estimation and plot their sample covariances, where the X-axis indicates the distance between spatial objects given by, say, the distance between centroids.

Sample covariance between objects $i$ and $j$:

$$c(h) = (y_i - \bar{y})(y_j - \bar{y})$$

$h$: the distance between objects $i$ and $j$: $c(h)$
3. Fit a theoretical covariance to the data by the least square method, and estimate the variance-covariance matrix $C$. 

4. Using the estimated $C$, estimate the parameter vector $\hat{\beta}$ by 

$$\hat{\beta} = (X'C^+X)^+X'C^+Y$$

5. Calculate the estimator of $Y$. 

6. Repeat the steps 2-5 until $\beta$ and $C$ converges to stable values (iterative process). Iterative estimation is often computationally expensive, but it is very useful when several parameters have to be estimated simultaneously.

Note

Application of the generalized regression model is not limited to spatial analysis; what makes the model different from the ordinary regression model is that it considers the correlation between probabilistic errors explicitly. Therefore, there are other methods to estimate the model that do not use the covariogram; in some cases the variance-covariance $C$ is given exogenously.

9.2.4 SEM: Spatial Error Model

General regression model is far more flexible than the ordinary regression model in that the correlation among error terms is considered explicitly. This model, however, is still based on one strong assumption - variances and covariances depends only on the distance between spatial objects. Though this assumption is indispensable in practical model estimation, it is not so realistic in the real world.

Model structure

The spatial error model relaxes this assumption; covariances among error terms is not restricted to a continuous function of the distance between spatial objects.

$Y$: Dependent variable
$X$: Independent variable
$U$: Structural error term
$W$: Distance matrix between spatial objects
$\rho$: Constant term
$\varepsilon$: Random error term

$$Y = X\beta + U$$
$$U = \rho W U + \varepsilon$$
Error terms

The spatial error model decomposes the probabilistic error term into structural part $U$ and random part $\varepsilon$.

The term $U$ indicates the probabilistic errors in the model that are correlated between spatial objects.

The term $\varepsilon$ is the independent probabilistic errors, that is, each error follows an identical probabilistic distribution independently.

The random error term $\varepsilon$ is assumed to satisfy

\[
E[\varepsilon] = 0 \\
V[\varepsilon] = E[\varepsilon \varepsilon^T] = \sigma^2 I_n
\]

Distance matrix between spatial objects

The distance matrix $W$ is written as

\[
W = \begin{bmatrix}
0 & d_{12} & d_{13} \\
d_{12} & 0 & d_{23} \\
d_{13} & d_{23} & 0
\end{bmatrix}
\]

where $d_{ij}$ is the distance between objects $i$ and $j$.

Meaning of the model

$Y = X\beta + U$

This equation is quite similar to the ordinary regression model; the dependent variable $Y$ is explained by a linear combination of independent variables $X$ and a probabilistic error term $U$.

The difference lies in the definition of error term $U$.

Structural error term $U$

$U = \rho WU + \varepsilon$

How is the structural error term $U$ determined? To answer this question, let us discuss in detail the term $U$.

The $i$th element of $U$ is expanded as

\[
u_i = \rho (d_{i1}u_1 + d_{i2}u_2 + \cdots + d_{i(i-1)}u_{i-1} + d_{i(i+1)}u_{i+1} + \cdots + d_{in}u_n) + \varepsilon_i
\]

(Note that the distance element $d_{ii}=0$). This indicates that the error $\nu_i$ is determined by the other errors $u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n$, and the random error $\varepsilon_i$.  


Therefore, this equation indicates that structural errors are correlated with each other, and that the correlation is determined by the distance matrix $W$.

The expectation of the error term $U$ is 0:

$$E[U] = E[(1 - \rho W)^{-1} E] = (1 - \rho W)^{-1} E[e] = 0$$

The variance-covariance matrix of the error term $U$ is thus given by

$$C = E[U U^T] = E[(1 - \rho W)^{-1} E^{T}] = (1 - \rho W)^{-1} E^{T} (1 - \rho W)^{-T} = \sigma^2 (1 - \rho W) (1 - \rho W)^T$$

Under this assumption, the likelihood is given by

$$L = \frac{1}{(2\pi)^{n/2} C^{1/2}} \exp \left[ -\frac{1}{2} (Y - X\beta)^T C^{-1} (Y - X\beta) \right]$$

The problem then becomes

$$\max_{\beta, \sigma, \rho} \left[ -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log C - \frac{1}{2} (Y - X\beta)^T C^{-1} (Y - X\beta) \right]$$

$$\approx \max_{\beta, \sigma, \rho} \left[ -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log C - \frac{1}{2} \sigma^2 (1 - \rho W) (1 - \rho W)^T \frac{1}{\sigma^2} (Y - X\beta) (1 - \rho W)(Y - X\beta) \right]$$

Unknown parameters to be estimated are $\beta$, $\sigma$, and $\rho$. To calculate those parameters we again use an iterative approach as we did in the generalized regression model.
Estimation procedure

1. Assume that the error term $U$ follows an identical probability distribution of mean 0 and variance $\sigma^2$ independently, that is, $U = \sigma^2 I_n$. Then apply the ordinary regression model to the data and estimate the parameter vector $\beta$. This gives the initial value of $\beta$.

2. Given the estimated $\beta$, calculate $\sigma$ and $\rho$ that maximize the likelihood:

$\max_{\sigma, \rho} \frac{1}{2} \log \det \sigma^2 (I - \rho W) + \frac{1}{n} (Y - X\hat{\beta})' \sigma^{-2} (I - \rho W) (Y - X\hat{\beta})$

Calculation of $\sigma$ and $\rho$ is a non-linear optimization problem; there are several approaches to this problem such as the steepest descent method, Newton method, and active set method.

3. Calculate $C$ using the estimated $\sigma$ and $\rho$:

$C = \sigma^2 (I - \rho W)' (I - \rho W)$

4. Assume that the error term $U$ follows the normal distribution with mean 0 and variance-covariance matrix $C$, and estimate $\beta$ by the generalized least square method:

$\hat{\beta} = (X'C^{-1}X)^{-1} X'C^{-1} Y$

5. Repeat the steps 2-4 until $\beta$, $\sigma$, and $\rho$ converges to stable values (iterative process).

9.2.5 SAR: Simultaneous AutoRegressive models

1: Simple model

The regression models discussed above explain the dependent variable of a spatial object by a set of independent variables of the same object.

The simultaneous autoregressive model, on the other hand, considers the interaction among spatial objects, that is, it the effect of dependent variables of spatial objects on those of other objects.

There are wide variety of simultaneous autoregressive models. Its simplest form is written as

$y_i = \sum_{j \neq i} g_{ij} y_j + \epsilon_i$

where

$y_i$: Dependent variable of region $i$

$g_{ij}$: Unknown parameter ($i, j = 1, \ldots, n$)
If we use the vector-matrix representation, it becomes:

\[ Y = GY + \epsilon \]

- **Y**: Dependent variable
- **G**: Unknown parameter matrix
- **\( \epsilon \)**: Random error term

This is the general form of the simple simultaneous autoregressive model. However, it is practically meaningless since the model cannot be estimated from observed data; the number of unknown parameters is larger than that of observations.

To reduce unknown parameters, we usually specify a function with a few parameters that represents the parameter matrix **G**.

A typical example is:

\[ g_{ij} = \frac{\gamma}{d_{ij}} \]

where \( d_{ij} \) is the distance between objects \( i \) and \( j \). The model becomes:

\[ Y = \sum_{ij} \frac{1}{d_{ij}} g_{ij} + \epsilon \]

If we write:

\[ W = \begin{bmatrix} 0 & d_{12} & d_{13} \\ d_{21} & 0 & \vdots \\ d_{31} & d_{32} & 0 \end{bmatrix} \]

the model is:

\[ Y = \gamma WY + \epsilon \]

Model estimation

The simultaneous model is estimated by the maximum likelihood method.

We assume that the error term \( \epsilon \) follows the identical normal distribution \( N(0, \sigma^2) \) independently.

Under this assumption, the likelihood is given by:

\[ L = \frac{1}{(2\pi \sigma^2)|1-yW|^2} \exp \left( -\frac{1}{2\sigma^2} (1-yW)'(1-yW) \right) \]

The problem then becomes:

\[ \max_{\gamma} \max_{\sigma^2} \left( -\frac{1}{2} \log \frac{1}{\sigma^2} (1-yW)'(1-yW) \right) \]
We have two unknown parameters, $\beta$ and $\gamma$. We can estimate them by solving the maximization problem using an optimization method such as the steepest descent method and the Newton method.

### 9.2.6 SAR: Simultaneous AutoRegressive models

#### 2: General model

The simple simultaneous autoregressive model considers the autocorrelation of dependent variables among spatial objects while the spatial error model focuses on the autocorrelation of probabilistic errors.

A more general form of the simultaneous autoregressive model treats both the autocorrelation of independent variables and probabilistic errors simultaneously in the model.

A more general form of the simultaneous autoregressive model is written as:

$$ Y = X\beta + GY + \varepsilon $$

- $X$: Independent variable
- $Y$: Dependent variable
- $\beta$: Unknown parameter vector
- $G$: Unknown parameter matrix
- $\varepsilon$: Random error term

To reduce the unknown parameters, we use a known weight matrix $W$ as we did in the previous section. The model then becomes:

$$ Y = X\beta + \gamma WY + \varepsilon $$

As a matter of fact, the spatial error model, which was discussed earlier, is a kind of simultaneous autoregressive model. The model is represented as:

$$ Y = X\beta + U $$

$$ U = \rho WU + \varepsilon $$

Substituting the second equation to the first one, we have:

$$ Y = X\beta + U $$

$$ = X\beta + \rho WU + \varepsilon $$

$$ = X\beta + \rho W(Y - X\beta) + \varepsilon $$

$$ = (1 - \rho W)X\beta + \rho WY + \varepsilon $$

It is quite similar to the general autoregressive model:

$$ Y = X\beta + \gamma WY + \varepsilon $$
This suggests that the general autoregressive model can be estimated by a procedure similar to that of the spatial error model; the model is estimated by the maximum likelihood method in which the ordinary and general least square methods are alternately used.

9.2.7 GLIM: Generalized Linear Model

So far we have discussed the numerical attributes of spatial objects as the dependent variable that are continuous and can take any value from \(-\infty\) to \(\infty\).

This is essential in the regression models since it permits normality assumption of the error term (the domain of the normal distribution is \((\infty, \infty)\)).

In spatial analysis, however, we often face variables that do not satisfy this assumption: population count data are positive integer variables; population density takes only positive values; land use ratio takes values from 0 to 1.

The problem consists of two parts: the problem of variable domain, and that of variable type.

The variable domain problem is relatively easy to solve; we transform the dependent variable so that it can take any value from \(-\infty\) to \(\infty\).

The variable type problem is not an easy problem; how do we treat count data including binary data?

The generalized linear model was developed to treat this problem.

If the dependent variable takes only the positive value, we apply the logarithmic transformation to the data:

\[ Y' = \log Y \]

If the dependent variable takes values from 0 to 1, we apply the logistic transformation to the data:

\[ Y' = \log \left( \frac{Y}{1-Y} \right) \quad \lambda \neq 0 \]

Box-Cox transformation? (Tango)
The ordinary regression model

\[ Y = X\beta + \varepsilon \]

means that the dependent variable \( Y \) is decomposed into the deterministic component \( X\beta \) and the probabilistic error \( \varepsilon \) which follows a probability distribution of mean 0.

Another interpretation of the ordinary regression model is that the dependent variable \( Y \) follows a probability distribution of mean \( X\beta \).

Let \( \mu \) be the mean value of the dependent variable \( Y \), that is, \( E[Y] = \mu \). The ordinary regression model is then represented as

\[ \mu = X\beta \]

The generalized linear model extends this representation of the ordinary regression model:

\[ g(\mu) = X\beta \]

where \( g() \) is called the link function.

The link function is a generalization of the variable transformation. The logarithmic transformation corresponds to the link function

\[ g(\mu) = \log \mu \]

The logistic transformation is represented by

\[ g(\mu) = \log \frac{\mu}{(1-\mu)} \]

If we use the link function

\[ g(\mu) = \mu \]

the generalized linear model is equivalent to the ordinary regression model.

The link function enables us to apply various types of transformations to the dependent variable \( Y \).

Besides the transformation of the dependent function, the generalized linear model has one advantage.

The ordinary regression model assumes that the dependent variable \( Y \) follows a probability distribution of mean \( X\beta \). Specifically, it assumes the normal distribution because it is necessary for testing the significance of the model.

The generalized linear model, on the other hand, assumes that the dependent variable \( Y \) follows a probabilistic distribution of the exponential family. The family includes the normal, binomial, and Poisson distributions, so this assumption gives wider flexibility than the ordinary regression model.

Obviously, if we use the normal distribution, the model is equivalent to the ordinary regression model.
When $Y$ is a positive integer variable such as population count, it is reasonable to assume the Poisson distribution:

$$Y \sim \text{Poisson}(\mu)$$

Considering the domain of $Y$ ($0 \leq Y$), we generally use the logarithmic transformation as the link function:

$$g(\mu) = \log \mu$$

This model is called the Poisson regression model.

When $Y$ is a binary variable, say, whether a region is urban or rural, the binomial distribution is appropriate:

$$Y \sim \text{Binomial}(\mu, 1)$$

Since the domain of $Y$ is $0 \leq Y \leq 1$, the logistic transformation is used as the link function:

$$g(\mu) = \log \left( \frac{\mu}{1-\mu} \right)$$

This model is called the logistic regression model.

Model estimation

To estimate the generalized linear model, we use the maximum likelihood method.

The process of model estimation is again iterative.

Estimation procedure

1. Apply the ordinary regression model to the data and estimate the parameter vector $\beta$. This gives the initial value of $\beta$.
2. Using the estimated $\beta$, compute the variance-covariance matrix of $Y$.
3. Estimate $\beta$ using the variance-covariance matrix of $Y$.
4. Repeat the steps 2-3 until $\beta$ converges to a stable value.

There are several commercial software packages such as GLIM that handle the generalized linear models flexibly.

9.2.7 Other regression-based models

There are still several important spatial models based on the regression model.

1. Geographically Weighted Regression Model (GWR)
2. Bidimensional regression model

Geographically Weighted Regression Model (GWR)

A kind of a regression model that explicitly considers the spatial heterogeneity of parameter values.
Bidimensional Regression Model

A linear regression model whose the independent and dependent variables are both two-dimensional vectors. Bidimensional regression model is used in cartogram, in order to fit a distorted map to a correct map.

\[
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix}
= \left( \begin{pmatrix}
x_1 & y_1 \\
x_2 & y_2 \\
\end{pmatrix}
\right)^T \left( \begin{pmatrix}
x_1 & y_1 \\
x_2 & y_2 \\
\end{pmatrix}
\right)
\]

Homework Q.9.1 (10 pts)

Show how the estimators of parameters $\beta$ and $\sigma^2$ are derived by the ordinary least square method in the ordinary regression model.

\[
\hat{\beta} = \left( X^T X \right)^{-1} X^T Y
\]

\[
\sigma^2 = \frac{1}{n-(k+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

Homework Q.9.2 (10 pts)

In regression model, parameter estimation is difficult if correlation exists among independent variables. This problem is called 'multicolinearity.'

1) How is the estimation difficult?
2) List some of the symptoms of multicolinearity.
3) How can we practically reduce or eliminate multicolinearity?