# 13

# *Regression Models for Spatially Autocorrelated Data*

## 13.1 Introduction

In this chapter, we discuss regression models specifically developed for processes described by spatially autocorrelated random variables. The effect of spatial autocorrelation, or apparent spatial autocorrelation, on regression models depends on how these spatial effects influence the data (Cliff and Ord, 1981, p. 141). Apparent autocorrelation may or may not be the result of real autocorrelation. Miron (1984) discusses three sources of real or apparent spatial autocorrelation: interaction, reaction, and misspecification. These are not mutually exclusive and may exist in any combination. We will discuss these in the context of a population of plants (for specificity, let us say they are oak trees in an oak woodland like that of Data Set 2) growing in a particular region. Suppose $Y_i$ represents a measure of plant productivity such as tree height or population density at location $(x_i, y_i)$, and that the population is sufficiently dense relative to the spatial scale that the productivity measure may be modeled as varying continuously with location. For purposes of discussion, suppose that we model tree productivity via a linear regression of the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \tag{13.1}$$

where $X_{i1}$ represents the amount of light available at location $i$, $X_{i2}$ represents the amount of available nutrients. In matrix notation, the ordinary least squares model is

$$Y = \beta X + \varepsilon. \tag{13.2}$$

We will set up models using artificial data to demonstrate the various effects of spatial autocorrelation.

The first source of autocorrelation we will consider is *interaction*. Negative autocorrelation may occur if trees in close proximity compete with each other for light and nutrients, so that relatively productive tree populations tend to inhibit the growth of other trees. Positive autocorrelation would occur if existing trees produce acorns that don't disperse very far, which in turn results in more trees in the vicinity. We will only consider positive autocorrelation, which seems to be much more common in ecology.

First, we generate a set of data that satisfies the model.

```
> set.seed(123)
> X1 <- rnorm(100)
> X2 <- rnorm(100)
> eps <- 0.01 * rnorm(100)
> b <- c(0.5,0.3)
> Y <- b[1]*X1 + b[2]*X2 + eps
> print(coef(lm(Y ~ X1 + X2)), digits = 2)
(Intercept)          X1          X2
     0.0014      0.4987      0.3002
```

If $Y$ is positively autocorrelated, then in place of Equation 13.2, the true model for its dependence on $X$ is

$$Y = \beta X + \rho WY + \varepsilon. \tag{13.3}$$

Here the term $\rho WY$ is the *spatial lag* (Section 3.5.2). We can generate a set of artificial data satisfying this model by employing the method of Haining (1990, p. 116) described in Section 3.5.3. Subtracting $\rho WY$ from both sides of Equation 13.3 and premultiplying the result by $(I - \rho W)^{-1}$ yields

$$Y = (I - \rho W)^{-1}(\beta X + \varepsilon). \tag{13.4}$$

Here is the code.

```
> rho <- 0.6
> library(spdep)
> nlist <- cell2nb(10, 10)
> IrWinv <- invIrM(nlist, rho)
> Y <- IrWinv %*% (b[1]*X1 + b[2]*X2 + eps)
```

Now let's try the regression again.

```
> print(coef(lm(Y ~ X1 + X2)), digits = 2)
(Intercept)          X1          X2
      0.021       0.542       0.324
```

The regression coefficients have all moved a bit away from their true values of 0, 0.5, and 0.3. The difference is not very great because $Y$ is not very large. Nevertheless, this example does illustrate the effect of positive interaction among the $Y$. The reason that omitting interactive spatial autocorrelation affects the regression coefficients can be seen by comparing Equations 13.2 and 13.3. Because Equation 13.2 does not include the lag term, when the data are fit to this model, some of the variability that would be assigned to this term if it were present is instead assigned to the regression coefficients. For example, the true marginal effect of nutrient level $X_2$, which is measured by $\beta_2$, will be incorrectly estimated because it will include some of the effect of the lag $WY$.

The second source of autocorrelation is called *reaction*. This would manifest itself in a plant population if nearby plants were reacting to the values of some external factor, such as soil nutrient content, that varies in space. If the primary influence on the process is one of reaction to the general ambient conditions, the explicit inclusion of spatial

autocorrelation terms may not be necessary or even appropriate. Instead, a simple regression model including the external factor, or inclusion of a trend model, may be appropriate. Miron (1984) discusses the potential effects of an omitted spatial influence on the parameters of a regression model. This effect can be visualized by supposing that the oak woodland we are modeling is a riparian area, and that the trees are affected by their distance to the river, either due to the soil properties or the availability of water or both. Replicating the discussion of Miron (1984) in an ecological context, the failure to include this effect means that there is a missing variable that we can denote $X_{i3}$, the distance from the river at location $i$. Instead of Equation 13.1, the model should be

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i. \tag{13.5}$$

Suppose, in addition, that $X_3$ is correlated with one of the other explanatory variables, say, with nutrient level $X_2$. In that case, the exclusion of distance from the river $X_3$ from the model may bias the coefficient $\beta_2$. The reason is again that the effect of $X_3$ has been "loaded" onto the explanatory variable $X_2$. We saw an example of this in Section 8.4.2 with the spatial variable *CoastDist*. In the analysis of the Sierra Nevada data carried out in the text, *CoastDist* did not have a strong effect, but if you did Exercise 8.12, you had the opportunity to show that in the Coast Range excluding *CoastDist* did impose a small bias on the other coefficients.

We will use the simulation to show the effects of excluding $X_3$ from the model when the process depends on this variable. In the model, the equation defining $X_3$ is a pure trend in the $x$ direction. Nutrient level $X_2$ also depends on distance from the river so that $X_2$ and $X_3$ are highly correlated, although there is not necessarily any causal relationship. We imbed the models in a function that can be used for Monte Carlo simulation.

```
> set.seed(123)
> coords <- expand.grid(1:10,10:1)
> b <- c(0.5,0.3,0.8)
> reg.mc <- function(coords, b, incl.X3){
+   X1 <- rnorm(100)
+   X2 <- coords[,1]+rnorm(100)
+   X3 <- coords[,1]+rnorm(100)
+   Y <- b[1]*X1 + b[2]*X2 + b[3]*X3 + rnorm(100)
+   {if (incl.X3 == T)
+      Y.lm <- lm(Y ~ X1 + X2 + X3)
+   else
+      Y.lm <- lm(Y ~ X1 + X2)}
+   return(coef(Y.lm))
+ }
```

If the third argument of the function `reg.mc(coords, b, incl.X3)` is `TRUE`, then $X_3$ is included in the model; otherwise, it is excluded. We first simulate the case in which $X_3$ is in the model.

```
> U <- replicate(1000,reg.mc(coords, b, T))
> print(rowMeans(U), digits = 2)
(Intercept)          X1          X2          X3
      0.014       0.499       0.303       0.796
```

The coefficients $\beta_1$ and $\beta_2$ are estimated without apparent bias.

In the second simulation, $X_3$ is not included in the model.

```
> U <- replicate(1000,reg.mc(coords, b, F))
> print(rowMeans(U), digits = 2)
(Intercept)          X1          X2
       0.46        0.49        1.02
```

Your numbers may be slightly different depending on the pseudorandom number generator. The estimate of $\beta_2$ is biased while that of $\beta_1$ is not. In this example, although $X_3$ is interpreted as a "spatial" variable, its role in the model is identical to that which it would play if it were interpreted as purely an attribute value, with no spatial connotation.

A second potential reaction effect of a spatial variable not included in the model is as follows. Suppose $X_3$ is not correlated with any of the other explanatory variables in the model. If $X_3$ is spatially autocorrelated and is not included in the model, then because its effect is loaded entirely into the error term, the result will be that the errors are autocorrelated. As we have seen in Section 3.3.2, in the presence of autocorrelated errors the estimates of the coefficients remain unbiased, but the error variance is inflated, so that the estimates of the coefficients become inefficient (i.e., needing more observations to achieve a conclusion). We can again demonstrate this effect through Monte Carlo simulation by modifying the data so that $X_1$, $X_2$, and $X_3$ are mutually uncorrelated and comparing simulations including and excluding $X_3$ in the model.

```
> library(spdep)
> set.seed(123)
> coords <- expand.grid(1:10,10:1)
> b <- c(0.5,0.3,0.8)
> lambda <- 0.6
> nlist <- cell2nb(10, 10)
> IrWinv <- invIrM(nlist, lambda)
> reg.mc <- function(coords, b, incl.X3){
+   X1 <- rnorm(100)
+   X2 <- rnorm(100)
+   X3 <- IrWinv %*% rnorm(100)
+   Y <- b[1]*X1 + b[2]*X2 + b[3]*X3 + rnorm(100)
+   {if (incl.X3 == T)
+      Y.lm <- lm(Y ~ X1 + X2 + X3)
+    else
+      Y.lm <- lm(Y ~ X1 + X2)}
+    return(coef(Y.lm))
+ }
```

First, we run a simulation in which $X_3$ is included in the model.

```
> U <- replicate(1000,reg.mc(coords, b, T))
> print(rowMeans(U), digits = 2)
(Intercept)          X1          X2          X3
     0.0088      0.4981      0.3029      0.7954
> print(apply(U, 1, var), digits = 2)
(Intercept)          X1          X2          X3
     0.0117      0.0105      0.0114      0.0084
```

The variable $X_3$ is autocorrelated. When it is included in the model, the variances of the coefficients are all about 0.01. Next, we exclude $X_3$ from the model and repeat the simulation.

```
> U <- replicate(1000,reg.mc(coords, b, F))
> print(rowMeans(U), digits = 2)
(Intercept)          X1          X2
   -0.00084     0.49357     0.29666
> print(apply(U, 1, var), digits = 2)
(Intercept)          X1          X2
      0.051       0.020       0.020
```

When $X_3$ is not included, the errors in the model become autocorrelated due to the autocorrelation of $X_3$. Unlike the previous example, because $X_2$ is independent of $X_3$ the estimate of $\beta_2$ is unaffected by removing $X_3$ from the model. The variance of the estimates of $\beta_1$ and $\beta_2$, however, are roughly doubled.

The third problem discussed by Miron (1984) is the effect of *model misspecification*. In this case, the measured autocorrelation is not due to interaction or reaction but to the incorrect form of the model. In particular, specification of homoscedastic errors when in fact the errors are heteroscedastic can lead to both biased estimates of the regression coefficients and indication of spatial autocorrelation when none really exists. To use Miron's term, this can "warp" the error terms of the model, creating a spatially heteroscedastic error, that is, non-constant variance (see Section 8.3). Suppose, for example, that in the tree growth model, nutrient level $X_2$ is spatially autocorrelated and that the error variance (i.e., the randomness associated with $X_2$) is an increasing function of $X_2$, but is not itself spatially autocorrelated (remember that it is spatial autocorrelation in the errors, not the in explanatory variables, normally introduces problems into regression results). This heteroscedasticity can induce apparent spatial dependency into the errors. We will demonstrate this with a simulation arranged so that $X_2$ has a spatial pattern similar to that displayed in Figure 3.9, with $\lambda = 0.8$. The error terms are uncorrelated, but because the error variance is a function of $X_2$ and high values of $X_2$ tend to be near other high values of $X_2$, a test for spatial autocorrelation of the residuals has a high Type I error rate. We use the function `lm.morantest()`, discussed in Section 13.2, to test for spatial autocorrelation of the residuals.

In the code below $X_2$ is spatially autocorrelated. The model for $Y_1$ is correctly specified, with constant variance errors, while the model for $Y_2$ is incorrectly specified, with heteroscedastic but not directly correlated errors.

```
> set.seed(123)
> coords <- expand.grid(1:10,10:1)
> b <- c(0.5,0.3)
> nlist <- cell2nb(10,10)
> IrWinv <- invIrM(nlist, 0.8)
> W <- nb2listw(nlist)
> reg.mc <- function(coords, b, W){
+   X1 <- rnorm(100)
+   X2 <- IrWinv %*% rnorm(100)
+   Y1 <- b[1]*X1 + b[2]*X2 + rnorm(100)
+   Y2 <- b[1]*X1 + b[2]*X2 + exp(1 + 2 * X2) * rnorm(100)
+ # Model correctly specified
```

```
+   Y1.lm <- lm(Y1 ~ X1 + X2)
+   t1 <- lm.morantest(Y1.lm, W)
+   C1 <- as.numeric(t1$p.value < 0.05)
+ # Model mis-specified, heteroscedastic errors
+   Y2.lm <- lm(Y2 ~ X1 + X2)
+   t2 <- lm.morantest(Y2.lm, W)
+   C2 <- as.numeric(t2$p.value < 0.05)
+   return(c(C1, C2))
+ }
> U <- replicate(1000,reg.mc(coords, b, W))
> rowMeans(U)
[1] 0.045 0.112
```

Significant spatial autocorrelation of the residuals is detected at the nominal rate of 5% in the correctly specified models, but it is detected in about 10% of the incorrectly specified ones.

Anselin and Griffith (1988) carry out a systematic investigation of the impacts of spatial effects on least squares regression models. In particular, they study the impact of autocorrelated errors on the Mallows $C_p$ (Section 9.1). Anselin and Griffith find that positively autocorrelated errors can artificially inflate the size of $C_p$. Anselin and Griffith also find that spatial autocorrelation of the errors can affect the outcome of tests for homogeneity of the residuals such as the Breusch-Pagan test (Kutner et al., 2005, p. 118). They find that autocorrelation of sufficient magnitude can increase the Type I error rates of certain tests, although the effect is not consistent across all types of tests.

In summary, models can be impacted by failure to account for spatial effects in regression model, either through the exclusion of a reaction to a spatially autocorrelated explanatory variable or by the presence of interactive spatial autocorrelation in the response variable or the error terms or both. They can also be impacted by misspecification of the model. Depending on the manner in which the effects of spatial effects enter into the system, these impacts can include biased estimates of the regression coefficients, biased estimates of the error variance, inefficient estimates of the regression coefficients, or spurious spatial autocorrelation.

Spatial autocorrelation often manifests itself in autocorrelation of the components of the error vector $\varepsilon$ in Equation 13.1. Section 13.2 contains a discussion of the detection of this type of autocorrelation. The field of spatial econometrics has generated a particularly useful approach to spatial regression, and much of the material in this chapter is motivated by that approach. Spatial autocorrelation is frequently incorporated into linear regression through one of two models, the spatial lag model or the spatial error model. Section 13.3 discusses these, and Section 13.4 describes a method for determining which is most appropriate for a particular data set. Section 13.5 describes the procedure for fitting the model. Section 13.6 describes an alternative model called the conditional autoregressive model, and Section 13.7 demonstrates the application of these models to field data.

## 13.2 Detecting Spatial Autocorrelation in a Regression Model

Consider the ordinary linear regression model, written in matrix notation (Appendix A.2) as

$$Y = X\beta + \varepsilon, \tag{13.6}$$

where $Y$ is a vector of $n$ response variables, $\beta$ is a vector of $p-1$ parameters, $X$ is an $n \times p$ design matrix, and $\varepsilon$ is a vector of $n$ error terms. The Gauss-Markov theorem states (Kutner et al., 2005, p. 18) that if the data conform to the model (Equation 13.6) and the components $\varepsilon_i$ of the error vector are independent, identically distributed random variables with mean zero and variance $\sigma^2$, then the method of least squares provides the best linear unbiased estimate of the parameters $\beta$. Spatial autocorrelation can disrupt least squares regression if the error terms $\varepsilon_i$ are spatially autocorrelated. One of the first questions to be addressed is, therefore, whether spatial autocorrelation exists among the error terms.

A number of tests for autocorrelation, prominent among them the use of Moran's $I$, are discussed in Chapter 4. Since the regression residuals are generally used as a model for the error terms, one might think that spatial autocorrelation could be detected by carrying out a Moran's $I$ test on these residuals. Unfortunately, things are not that simple (Cliff and Ord, 1981, p. 200; Miron, 1984). Suppose a least squares regression is carried out on a data set $Y = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}'$ resulting in a regression model

$$Y = Xb + e. \tag{13.7}$$

The regression coefficients $b$ satisfy (Kutner et al., 2005, p. 200)

$$b = (X'X)^{-1}X'Y, \tag{13.8}$$

and therefore the regression residuals satisfy

$$\begin{aligned} e &= Y - Xb \\ &= (I - X(X'X)^{-1}X')Y \\ &= MY \end{aligned} \tag{13.9}$$

where $M = I - X(X'X)^{-1}X'$. The formula for the variance–covariance matrix of $e$ is then $\text{var}\{e\} = \text{var}\{MY\}$ (Kutner et al., 2005, p. 196). Therefore,

$$\begin{aligned} \text{var}\{e\} &= \text{var}\{MY\} \\ &= M\,\text{var}\{Y\}M' \\ &= M\sigma^2 IM' \\ &= \sigma^2 MM' \end{aligned} \tag{13.10}$$

It turns out (Kutner et al., 2005, p. 204) that the matrix $M$ is *idempotent*, that is, it satisfies $MM' = M$. We get the final formula

$$\text{var}\{e\} = \sigma^2 M. \tag{13.11}$$

Since $M$ has nonzero off-diagonal terms, it follows that $\text{cov}\{e_i, e_j\} \neq 0$ in general. Therefore, the regression residuals are correlated, and a test using Moran's $I$ on these residuals may indicate autocorrelation even if the error terms $\varepsilon_i$ are themselves uncorrelated. It is worth noting that one could have determined that the residuals $e_i$ are correlated without actually computing Equation 13.11. Indeed, they must satisfy (Kutner et al., 2005, p. 23) $\Sigma_i e_i = 0$, and given any $n-1$ of them, one can compute the last. Therefore, they must be correlated with each other. To say that the residuals are correlated does not mean that they are *spatially*

autocorrelated, but the possibility exists that they are. Cliff and Ord (1981, p. 200) have derived formulas for the mean and variance of Moran's $I$ for regression residuals. Any test of autocorrelation of ordinary least squares residuals must be carried out using these formulas rather than those given in Chapter 4.

The `spdep` package includes a function `lm.morantest()` that takes into account the correlation of residuals in carrying out the test for autocorrelation. This test was used in Section 13.1 to detect apparent spatial autocorrelation in a misspecified model.

As an illustration of the test of a real model for spatially autocorrelated residuals, we use the linear model `model.5` for Field 4.1 developed in Section 9.3.

```
> model.5 <- lm(Yield ~ Clay + SoilP + I(Clay*SoilP) + Weeds,
+    data = data.Set4.1)
> library(spdep)
> coordinates(data.Set4.1) <- c("Easting", "Northing")
> nlist <- dnearneigh(data.Set4.1, d1 = 0, d2 = 61)
> W <- nb2listw(nlist, style = "W")
> lm.morantest(model.5, W)
        Global Moran's I for regression residuals
data :
model: lm(formula = Yield ~ Clay + SoilP + I(Clay * SoilP) + Weeds, data
= data.Set4.1)
weights: W
Moran I statistic standard deviate = 1.6864, p-value = 0.04586
alternative hypothesis: greater
sample estimates:
Observed Moran's I        Expectation           Variance
       0.091175267       -0.040897372        0.006133274
```

The resulting $p$ value supports the conclusion that the errors in this model are spatially autocorrelated. Therefore, to safely use this model we must determine the most appropriate form in which to include the effects of spatial autocorrelation. This is the subject of the next section.

## 13.3 Models for Spatial Processes

### 13.3.1 The Spatial Lag Model

In Chapter 3, it was pointed out that a close relationship exists between spatial processes and time series. Two spatial models were briefly discussed, the *spatial lag model* and the *spatial error model*. In this section, we consider these and other models in greater detail. We begin our discussion with the spatial lag model. In time series, a fundamental model is the so-called *first-order autoregressive model*, (Box and Jenkins, 1976, p. 51), which has the form

$$Y_i = \rho Y_{i-1} + \varepsilon_i,\ i = 1, 2, \ldots \tag{13.12}$$

The equation in matrix notation for the analogous spatial model, which describes a spatially lagged random variable with zero mean and no explanatory variables, has the form

$$Y = \rho WY + \varepsilon. \tag{13.13}$$

As discussed in Chapter 3, the quantity $WY$ is called the *spatial lag* by analogy with the time lag $Y_{i-1}$ in a time series model. The interpretation of model (Equation 13.13) is that the value of the process $Y$ at one location is directly associated with the values of the process at nearby locations. In the example discussed in Section 13.1, $Y$ represents a measure of oak tree productivity, and high productivity at one location is associated with high productivity at other nearby locations. It is important to remember that nothing in the model says that high productivity at one location actually *causes* high productivity at nearby locations.

The spatial lag model, which is also sometimes called the *simultaneous autoregressive model*, was initially studied by Whittle (1954) in a paper that played a pivotal role in initiating the study of spatial processes. Whittle considered (in the terminology of Chapter 3) a binary rook's case spatial relationship. Anselin (1992, p. 34) describes higher-order contiguity models (cf. the discussion of the correlogram in Section 4.6), but in this text we will only consider first-order contiguity. Inverting Equation 13.13 leads to

$$Y = (I - \rho W)^{-1}\varepsilon, \tag{13.14}$$

Equations of this form have been used repeatedly throughout the book to generate spatially autocorrelated random variables.

In the analysis of the mixed model in Chapter 12, we made an assumption about the variance of $Y$, and therefore of the residuals, that enabled a model of the autocorrelation structure to be constructed (Equations 12.8 and 12.9). In the analysis of the autoregressive models of this chapter, the form of the variance of $Y$ is enforced by the model. From Equation 13.14 we have

$$\begin{aligned}
\text{var}\{Y\} &= E\{YY'\} \\
&= E\{(I - \rho W)^{-1}\varepsilon\varepsilon'[(I - \rho W)^{-1}]'\} \\
&= (I - \rho W)^{-1}E\{\varepsilon\varepsilon^T\}(I - \rho W')^{-1} \\
&= (I - \rho W)^{-1}\sigma^2 I(I - \rho W')^{-1},
\end{aligned} \tag{13.15}$$

where in going from the second to the third equation we have used the fact that the transpose of the inverse equals the inverse of the transpose (Appendix A.1). As with the mixed model, the variance of the autoregressive model is made up of a matrix product of terms that describe the spatial structure with a term that describes the magnitude of the variance. However, the roles of the terms in the last of Equations 13.15 is the reverse of that of the corresponding matrices in Equation 12.9.

It is evident that if we have a set of explanatory variables to incorporate into a spatial regression model, then this can be accomplished by augmenting the spatial lag model of Equation 13.13. Adding a term $X\beta$ to the spatial lag model of Equation 13.13 yields

$$Y = \rho WY + X\beta + \varepsilon. \tag{13.16}$$

This is sometimes called the mixed regressive–spatial auto regressive model (Anselin, 1988, p. 35), or sometimes simply the *spatial autoregressive model*. We will still use the term *spatial lag model* (Anselin, 1992, p. 48). In other words, we regard the model of Equation 13.13 as simply a special case of Equation 13.16 in which $\beta = 0$.

The model of Equation 13.16 can be interpreted in three different ways (Anselin, 1992, p. 36). In the first interpretation, one is interested in the model of the spatial process for

its own sake. A common (non-ecological) example of this is the study of the spread of technological innovations. For example, Anselin (1992, p. 57) studies the state-by-state spread of the adoption of home freezers in the United States. In this case, the proper specification of the spatial weights matrix $W$ and the estimation of the parameter $\rho$ become important in their own right as indicators of the nature and strength of spatial interaction.

A second interpretation is obtained by bringing the first term on the right-hand side of Equation 13.16 to the left side, resulting in

$$(I - \rho W)Y = X\beta + \varepsilon. \tag{13.17}$$

In this formulation, one is primarily interested in the influence of the predictor variables $X$ on the response variable after controlling for the effect of spatial autocorrelation. In this context, the quantity $(I - \rho W)$ is sometimes called the *spatial filter*.

A third interpretation is obtained by multiplying both sides of Equation 13.16 by $(I - \rho W)^{-1}$ and taking the expectation of both sides. This yields

$$\begin{aligned} E\{Y\} &= E\{(I - \rho W)^{-1} X\beta\} + E\{(I - \rho W)^{-1}\varepsilon\} \\ &= E\{(I - \rho W)^{-1} X\beta\}. \end{aligned} \tag{13.18}$$

since $E\{\varepsilon\} = 0$. In this interpretation, one focuses on the nonlinear effect of the spatial autocorrelation on the expected values of the components of $Y$.

### 13.3.2 The Spatial Error Model

In Chapter 3, we introduced the autoregressive time series model (Equation 3.13)

$$Y_i - \mu = \lambda(Y_{i-1} - \mu) + \varepsilon_i,\ i = 1, 2, \ldots \tag{13.19}$$

which (see Exercise 13.1) can be written as (Equation 3.12)

$$\begin{aligned} Y_i &= \mu + \eta_i \\ \eta_i &= \lambda\eta_{i-1} + \varepsilon_i,\ i = 1, 2, \ldots \end{aligned} \tag{13.20}$$

Note that in this case the lagged quantity is the *error* rather than the response variable. The spatial version of this model, including explanatory variables, is

$$\begin{aligned} Y &= X\beta + \eta \\ \eta &= \lambda W\eta + \varepsilon. \end{aligned} \tag{13.21}$$

This formulation is called the *spatial error model*. By tradition, the symbol $\rho$ is used with the spatial lag model, and $\lambda$ is used with the spatial error model. The first equation of Equations 13.21 is that of an ordinary linear regression model, but because of the second equation, the errors have a correlated structure. Spatial autocorrelation of the form (Equation 13.21) is generally considered a nuisance (Anselin, 1992, p. 38), because the primary interest is in the relationship between the explanatory variables $X$ and the response variable $Y$, and the effect of spatial autocorrelation is to prevent the data from satisfying the assumptions of ordinary least squares regression.

There is an algebraic relationship between the spatial error model and the spatial lag model (see also Exercise 13.1 for the time-series analogy). Indeed, subtracting $\lambda WY$ from both sides of the first of Equations 13.21 yields

$$Y - \lambda WY = X\beta - \lambda WY + \eta. \tag{13.22}$$

Substituting $X\beta + \eta$ for $Y$ and $-\lambda W\eta + \eta$ for $\varepsilon$ from Equations 13.21 gives

$$\begin{aligned} Y - \lambda WY &= X\beta - \lambda WY + \eta \\ &= X\beta - \lambda W(X\beta + \eta) + \eta \\ &= X\beta - \lambda WX\beta + (-\lambda W\eta + \eta) \\ &= X\beta - \lambda WX\beta + \varepsilon. \end{aligned} \tag{13.23}$$

Thus, the pair of equations (Equation 13.21) defining the spatial error model is equivalent to the single autoregressive model

$$Y - \lambda WY = X\beta - \lambda WX\beta + \varepsilon, \tag{13.24}$$

in which the spatial lag operator is applied to the explanatory variables as well as the response variable. Since the fixed term $X\beta$ in Equation 13.24 does not contribute to the variance of $Y$, we can set it equal to zero and see that the expression for the variance of the spatial error model is the same as that of the spatial lag model except that $\lambda$ is substituted for $\rho$.

It follows that if one knew the value of $\lambda$ one could define variables

$$\begin{aligned} \tilde{Y} &= Y - \lambda WY \\ \tilde{X} &= X - \lambda WX, \end{aligned} \tag{13.25}$$

and carry out a regression of $\tilde{Y}$ on $\tilde{X}$. This process is implemented in generalized least squares, which is discussed in Section 12.5 (Anselin, 1992, p. 44; Pinhiero and Bates, 2000, p. 203). Since the value of $\lambda$ is not known, it must be estimated by a method such as maximum likelihood. One can write a spatial autoregressive moving average model as (Anselin, 1992, p. 35)

$$Y = \rho WY + X\beta + \lambda W\varepsilon + \varepsilon, \tag{13.26}$$

and this can be generalized to higher orders as well. However, the theory of these models is not as well developed as is the theory for the spatial error and the spatial lag model, and we will not consider the possibility of fitting data to a model of the form (Equation 13.26). See LeSage and Pace (2009) for a discussion of the model.

## 13.4 Determining the Appropriate Regression Model

### 13.4.1 Formulation of the Problem

Suppose spatial autocorrelation has been detected in the residuals of a linear regression model, as it was, for example, in Section 13.2 for the model of yield in Field 1 of Data Set 4. The first step in augmenting the model to take this autocorrelation into account is

generally to determine whether the spatial lag model (Equation 13.16) or the spatial error model (Equation 13.21) provides a better fit to the data. Anselin et al. (1996) formulate the general problem of distinguishing between models (Equation 13.16) and (Equation 13.21) by combining them into a single model of the form

$$\begin{aligned} Y &= \rho W_1 Y + X\beta + \eta \\ \eta &= \lambda W_2 \eta + \varepsilon. \end{aligned} \tag{13.27}$$

where $\varepsilon \sim N(0, \sigma^2 I)$, and $W_1$ and $W_2$ are row-normalized spatial weights matrices. In order for the model to be identifiable, either $W_1$ cannot equal $W_2$ or the matrix $X$ cannot be the trivial matrix consisting of a vector of ones (Anselin et al., 1996, p. 81). We will assume that $X$ does indeed contain at least one true explanatory variable. The problem of distinguishing between the two models then becomes one of testing two separate hypotheses. The first is a test of $H_0 : \rho = 0$ against $H_a : \rho \neq 0$, and the second is a test of $H_0 : \lambda = 0$ against $H_a : \lambda \neq 0$. Anselin et al. (1996) carry out the test using an alternative to the likelihood ratio test called the *Lagrange multiplier test*. The next subsection contains a description of this test.

### 13.4.2 The Lagrange Multiplier Test

Suppose we have a statistical model with a parameter $\theta$ and we wish to test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_a : \theta \neq \theta_0$ using the method of maximum likelihood. As discussed in Section 12.2, one way to do this is to use the likelihood ratio test. If $l_F$ is the value of the log likelihood function $l(\theta \mid z)$ at maximum likelihood estimate when $\theta$ is not restricted to the value $\theta = \theta_0$ (the full model) and $l_R$ is the value of the log likelihood function when $\theta$ is restricted to $\theta = \theta_0$ (the restricted model), then the likelihood ratio test involves computing the statistic $G = -2\log(L_F / L_R) = 2(l_F - l_R)$ (Equation 12.4) and comparing its value with that of a chi-square distribution. For complex likelihood functions like those that arise from Equation 13.27, however, the likelihood function can be prohibitively difficult to compute, and therefore one would like to find a test involving a simpler function. One such test is the Lagrange multiplier test, also known as the score test.

The method of Lagrange multipliers is described in Appendix A.4. It is applied to the maximum likelihood test described in the first paragraph of this section as follows (Theil, 1971, p. 527; Fox, 1997, p. 570). Let $l(\theta|Y)$ be the log likelihood function, and let $\theta_{ML}$ be the maximum likelihood estimator of $\theta$. Consider the constrained maximization problem of maximizing $l(\theta \mid Y)$ subject to the constraint $\theta = \theta_{ML}$. Formulating this as a Lagrange multiplier problem implies finding the maximum of $l(\theta \mid Y) + \gamma(\theta_{ML} - \theta)$. Applying the Lagrange multiplier equations (Equation A.46) yields

$$\begin{aligned} \frac{\partial l}{\partial Y} - \gamma\theta &= 0 \\ \frac{\partial l}{\partial \gamma} = \theta_{ML} - \theta &= 0. \end{aligned} \tag{13.28}$$

Thus, the value of the Lagrange multiplier $\gamma$ is the slope $\partial l / \partial Y$ of the log likelihood function. The larger this slope at a particular value of $\theta$, the further this value of $\theta$ is from $\theta_{ML}$.

The slope of the log likelihood function at the value $\theta$ is called the *score* and is denoted $S(\theta)$. This does not appear to have simplified the problem, since we still must compute the derivative of the log likelihood. It turns out, however, that our effort does lead to a simplification. Anselin (1988a) developed a way to test hypotheses involving the model of Equation 13.27 using the Lagrange multiplier test, and Anselin et al. (1996) developed a simpler and more robust form of this test that functions equally well. Anselin and Rey (1991) compared the Lagrange multiplier test applied to model (Equation 13.27) with a hypothesis test based on the Moran's $I$ statistic and concluded that the Lagrange multiplier test is more appropriate, and this conclusion was further supported by the work of Anselin et al. (1996).

Both the original Lagrange multiplier test of Anselin (1988a) and the robust version of Anselin et al. (1996) are available in the function `lm.LMtests()` in the package `spdep`. The tests are carried out on the model

$$\begin{aligned} Y &= \rho WY + X\beta + \eta \\ \eta &= \lambda W\eta + \varepsilon. \end{aligned} \tag{13.29}$$

This is the same as the model of equations (Equation 13.27) except that $W_1 = W_2 = W$. This is acceptable since we have specified that $X$ is nontrivial. The tests are successively that $\lambda = 0$ and that $\rho = 0$. If either of these tests returns a nonsignificant $p$ value, then we cannot reject the null hypothesis that the corresponding autocorrelation coefficient is zero.

To implement the test, we first generate the spatial lag (Equation 13.16) and spatial error (Equation 13.21) models for the regression problem. Then we call the function `lm.LMtests()` with the argument `tests = "all"`, which runs all of the tests on the two models.

```
> model.5 <- lm(Yield ~ Clay + SoilP + I(Clay*SoilP) + Weeds,
>    data = data.Set4.1)
> library(spdep)
> coordinates(data.Set4.1) <- c("Easting", "Northing")
> nlist <- dnearneigh(data.Set4.1, d1 = 0, d2 = 61)
> lm.LMtests(model.5, W, test = "all")
        Lagrange multiplier diagnostics for spatial dependence
LMerr = 1.2303, df = 1, p-value = 0.2673
LMlag = 10.9906, df = 1, p-value = 0.0009157
RLMerr = 8.1514, df = 1, p-value = 0.004303
RLMlag = 17.9116, df = 1, p-value = 2.314e-05
SARMA = 19.142, df = 2, p-value = 6.972e-05
```

Most of the output of the function is suppressed; only the relevant $p$ values are shown. The $p$ values indicate a preference for the spatial lag model (Equation 13.16). The ecological interpretation of this result is that the crop yield values are directly associated with each other, as opposed to being associated with unmeasured, spatially autocorrelated processes that are loaded into the error. This may seem a bit counterintuitive and warrants further discussion, which we give at the end of Section 13.7. The SARMA, or spatially autoregressive moving average, model is also not rejected. This model is not covered here. In the next section, we will develop the methods necessary to fit both a spatial lag model and a spatial error model to these data.

## 13.5 Fitting the Spatial Lag and Spatial Error Models

In this section we discuss the problem of estimating the coefficients of the spatial lag model (Equation 13.16) and the spatial error model (Equation 13.21). The fitting of both models involves a similar procedure, and so we only discuss the spatial error model. The mathematics of this section is fairly complex and need not be totally mastered on first reading.

To begin, we review the solution by maximum likelihood of the simple linear regression problem

$$Y = X\beta + \varepsilon, \tag{13.30}$$

where $\varepsilon$ is a vector with elements $\varepsilon_i \sim N(0,\sigma^2)$ and the $\varepsilon_i$ are mutually independent (i.e., there is no spatial autocorrelation). Assume that the explanatory variables $X_i$ are mathematical variables as defined in Appendix A.2. The case in which the $X_i$ are random variables requires a longer argument to reach the same conclusion.

For notational simplicity, we will consider the case in which $p = 2$. The problem we must resolve is that when we compute the likelihood function $L(\beta_0, \beta_1, \sigma^2 \mid Y)$, we are computing the likelihood in which the vector $Y$, with components consisting of the random variables $Y_i$, is the parameter, but the distribution we know is that of the random variables $\varepsilon_i$. Specifically, the $\varepsilon_i$ satisfy $\varepsilon_i \sim N(0,\sigma^2)$. Since the $\varepsilon_i$ are the only random variables in Equation 13.30, the $Y_i$ have the same distribution as the $\varepsilon_i$, and therefore the $Y_i$ are also independent and normally distributed with variance $\sigma^2$. The fact that the $Y_i$ are independent permits us to write the likelihood as a product of normal density functions, but if they were not independent, then in order to write the likelihood as a product we would have to transform the variables from the $Y_i$ into the $\varepsilon_i$. In reality, we are also doing a transformation when the errors $\varepsilon_i$ are independent, but this transformation is just the identity $Y = I\varepsilon = \varepsilon$, so we can ignore it. When we deal with autocorrelated errors, however, the transformation will not be the identity and we will not be able to ignore it.

When the $\varepsilon_i$ are independent the likelihood function for Equation 13.30 is (Kutner et al., 2005, p. 27)

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2 \mid Y) &= \prod \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[ -\frac{1}{2}\left( \frac{Y_i - \beta_0 - \beta_1 X_i}{\sigma} \right)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \right). \end{aligned} \tag{13.31}$$

Computing $l(\beta_0, \beta_1, \sigma^2 \mid Y) = \log L(\beta_0, \beta_1, \sigma^2 \mid Y)$ and differentiating yields

$$\begin{aligned} \frac{\partial l}{\beta_0} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial l}{\partial \beta_1} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{2} (Y_i - \beta_0 - \beta_1 X_i)^2. \end{aligned} \tag{13.32}$$

Solving these yields

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

$$\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{13.33}$$

$$\frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2}{n} = \hat{\sigma}^2.$$

The first two equations are identical to the normal equations of ordinary least squares regression (Kutner et al., 2005, p. 17), and the third equation is a biased estimator of $\sigma^2$.

Now consider the autoregressive case, specifically, the spatial error model given by Equation 13.29. Analogous to Equations 13.30 and 13.31, we must compute a likelihood function $L(\beta, \sigma^2, \lambda \mid Y)$, where $\beta$ is the vector of regression parameters. However, $Y$ now has the same probability distribution as $\eta$ rather than $\varepsilon$, and whereas the $\varepsilon$ are independent, the $\eta$ are not. Moreover, in order to determine the probability distribution of the $\eta$ we need to know $\lambda$, but $\lambda$ is one of the parameters we are trying to estimate. Therefore, in order to compute the likelihood function, we must use the relationship $\eta = \lambda W\eta + \varepsilon$ to re-express in terms of $\eta$ the probability density that we currently have in terms of $\varepsilon$. In doing so, we have to be careful to preserve properties of the likelihood as a probability density function (Appendix A.5.1).

In Appendix A.6, it is shown that the equation for the likelihood $L$, expressed in terms of $\eta$, is

$$L(\beta, \lambda, \sigma^2 \mid \eta) = |\det A| \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(A\eta)'A\eta\right). \tag{13.34}$$

where the matrix $A$ is defined by

$$\varepsilon = (I - \lambda W)\eta \equiv A\eta. \tag{13.35}$$

Taking logarithms and expressing the equation $\eta = Y - X\beta$ yields

$$l(\beta, \lambda, \sigma^2 \mid Y) = \log|\det A| - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta\right) + C \tag{13.36}$$

where $C$ is a constant incorporating all the terms that do not depend on the parameters of the model. In order to obtain maximum likelihood estimates of the parameters, one must differentiate Equation 13.36 with respect to these parameters and set these partial derivatives equal to zero.

There are a number of different ways to maximize the log likelihood (Equation 13.36). The method described by Upton and Fingleton (1985, p. 285) provides a good example. If $\lambda$ were known, it would be a straightforward if somewhat messy task to differentiate (Equation 13.36) and solve the resulting equations. Indeed, differentiating $l$ with respect to $\beta$ and $\sigma^2$ yields

$$\frac{\partial l}{\partial \beta} = -\frac{1}{2\sigma^2}\left(-2X'A'AY + 2X'A'AX\beta\right)$$
$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}(Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta) \tag{13.37}$$

Setting these equal to zero and solving yields

$$\hat{\beta} = (X'A'AX)^{-1}X'A'AY$$
$$\hat{\sigma}^2 = \frac{1}{n}(Y'A'AY - 2\beta'X'A'AY + \beta'X'A'AX\beta) \tag{13.38}$$

The matrix depends on $\lambda$, which we do not know. If we did, we could compute $A$ and substitute the second of Equations 13.37 into Equation 13.36 to get

$$l(\beta, \lambda, \sigma^2 \mid Y) = \log|\det A| - \frac{n}{2}\log\hat{\sigma}^2 + C \tag{13.39}$$

We do not know $\lambda$, however, so we must also estimate its value. Since $A = I - \lambda W$, $\det A$ is an $n$th degree polynomial in $\lambda$, which considerably complicates things (Cliff and Ord, 1981, p. 155). Usually one assumes a value of $\lambda$ and then uses a numerical scheme analogous to the Newton-Raphson method described in Appendix A.5.2 to find the zeroes of the derivatives in Equations 13.37, and then one iterates this with solutions of Equation 13.39 until a maximum of $l$ is reached. Most software packages make use of some variant of this approach. We will postpone the application of the theory to data until after the next section, in which we discuss an alternative model for spatial autoregression.

## 13.6 The Conditional Autoregressive Model

The *conditional autoregressive*, or CAR, model is based on an alternative description of the probabilistic structure of the data. The models discussed in Section 13.3 describe the *joint* probability density functions of the $Y_i$. Although the terminology is not completely consistent, many authors refer to models such as these as *simultaneous autoregressive* or SAR models. The CAR model is constructed by specifying the *conditional* probability density of one of the $Y_i$ given all of the others. To introduce the CAR formalism, we return to the simple time series model $Y_i - \mu = \rho(Y_{i-1} - \mu) + \varepsilon_i$, $i = 1, 2, \ldots$ given in Equation 13.12 for the special case $\mu = 0$. This model could also be expressed as (Cliff and Ord, 1981, p. 146)

$$E\{Y_i \mid Y_{i-1}, Y_{i-2}, \ldots\} = \mu + \rho(Y_{i-1} - \mu),$$
$$\text{var}\{Y_i \mid Y_{i-1}, Y_{i-2}, \ldots\} = \sigma^2. \tag{13.40}$$

Equations 13.40 convey the idea that the properties of $Y_i$ are conditional on those of the preceding values in the time series. In the case of a time series, the formulations, Equation 13.40 and Equation 13.12, are equivalent, but in the case of spatial models they are not (Cliff and Ord, 1981, p. 146). Nevertheless, the CAR model can be developed by analogy to the conditional time series formulation (Equation 13.40). Specifically, let $Y_{-i}$ denote the set of all members of the vector $Y$ *except* $Y_i$. Then we can write the CAR formulation as (Waller and Gotway, 2004, p. 371)

$$\begin{aligned} E\{Y_i \mid Y_{-i}\} &= X\beta + \sum_{j \neq i} c_{ij}(Y_j - X\beta), \\ \text{var}\{Y_i \mid Y_{-i}\} &= \sigma^2, \end{aligned} \tag{13.41}$$

where $c_{ij} = c_{ji}$ and the conditional expectations are assumed to be normally distributed. Waller and Gotway (2004, p. 371) provide a nice analogy of this process in a classroom setting. Imagine a classroom in which the desks are arranged in a set of rows and columns. The instructor hands each student a slip of paper that only he or she can see and that contains the value of some quantity, say, temperature, measured at the location of that student's desk. The instructor then tells the students to raise their hands to a level that indicates the value of that quantity. The students are aware that the temperature of one desk would be close to the temperature at a nearby desk, so they surreptitiously peek around and see how high their neighbors are raising their hands, and each student adjusts his or her own hand accordingly. The students are thus conditioning their response on the values of those of their neighbors.

Equations 13.41 can easily be generalized to the case $\text{var}\{Y_i \mid Y_{-i}\} = \sigma_i^2$, that is, to the case in which the variances are not identical. We will not make this generalization here; the interested reader is referred to Waller and Gotway (2004, p. 371). There (see also Cliff and Ord, 1981, p. 180) it is shown that for fixed $\sigma^2$ the joint distribution of the $Y_i$ is multivariate normal with mean $X\beta$ and variance $\Sigma_Y = \sigma^2(I - C)^{-1}$, where $C$ is the symmetric matrix whose elements are the $c_{ij}$.

In order to ensure a well-specified probability structure, a set of conditions contained in a theorem known as the Hammersley-Clifford theorem must be satisfied. Haining (1990, p. 87) provides a discussion of these conditions, but for our purposes it is sufficient to say that the conditions are satisfied by any practical example that we will consider (Waller and Gotway, 2004, p. 372).

It turns out that any SAR model of the form (Equation 13.21) may be expressed as a CAR model where $C = W + W' - WW'$. However, CAR models have different properties and applications from SAR models and should be considered differently (Ripley, 1981, p. 88). One advantage of CAR models is that they may be solved using ordinary least squares (Haining, 1990, p. 130). A disadvantage of the CAR model is that it may happen that in order to ensure a valid least-squares estimate, part of the data may have to be left unused in the estimation (Upton and Fingleton, 1985, p. 363; Haining, 1990, p. 133). This latter problem, however, should not affect models constructed with a first-degree nearest-neighbor spatial weights matrix. From the perspective of this book, perhaps the most significant aspect of the CAR formalism is that it fits well into a Bayesian analysis, which will be covered in Chapter 14.

## 13.7 Application of Simultaneous Autoregressive and Conditional Autoregressive Models to Field Data

### 13.7.1 Fitting the Data

We will continue to work with `model.5` for the data of Field 1 of Data Set 4. As a reminder, the model is

$$Yield_i = \beta_0 + \beta_1 Clay_i + \beta_2 SoilP_i + \beta_3 Clay_i \times SoilP_i + \beta_4 Weeds_i + \varepsilon_i \tag{13.42}$$

The results in Section 13.4.2 lead us to believe that the errors $\varepsilon_i$ in this model may display significant spatial autocorrelation.

Our analysis follows those of Bivand et al. (2013b, Ch. 10), Waller and Gotway (2004, Ch. 9), and Schabenberger and Gotway (2005, Ch. 6). Before we begin the analysis, we must make one very important modification. Fitting the data to a spatial lag or spatial error model involves computing the determinant of the matrix $A$ in Equation 13.39, which is a *Jacobian* matrix (Appendix A.6). It turns out that if the ranges of the variables in the model are very different, then the numerical methods involved in the computation may fail (Bivand et al., 2013b, p. 292). For this reason, before beginning the analysis we will center and scale the variable *Yield*, which will reduce the size of the regression coefficients and make them more uniform. As usual, the data are loaded into R using the code in Appendix B.4. The data frame `data.Set4.1` holds the contents of *set4.196sample.csv*, and the data frame `data.Yield4.1idw` holds the yield monitor values interpolated to the 86 sample points.

```
> data.Set4.1$Yield <- scale(data.Yield4.1idw$Yield)
```

Here is a portion of the output of the function `summary()` for `model.5` with the scaled yield.

```
> formula.5 <- as.formula("Yield ~ Clay + SoilP +
+    I(Clay*SoilP) + Weeds")
> model.5 <- lm(formula.5, data = data.Set4.1)
> summary(model.5)
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.318959   1.075379   8.666 3.64e-13 ***
Clay            -0.229174   0.026715  -8.578 5.42e-13 ***
SoilP           -0.628251   0.144961  -4.334 4.17e-05 ***
I(Clay * SoilP)  0.016974   0.003724   4.558 1.81e-05 ***
Weeds           -0.275797   0.050309  -5.482 4.64e-07 ***
```

Since heteroscedasticity of errors can lead to problems with spatial data, we will check for it using the Levene test and the Breusch-Pagan test (Kutner et al., 2005, p. 116). The former is available in the `car` package (Fox and Weisberg, 2011), and the latter in the `lmtest` package (Zeileis and Hothorn, 2002).

```
> library(car)
> library(lmtest)
> bptest(model.5)
        studentized Breusch-Pagan test
data:  model.5
BP = 8.614, df = 4, p-value = 0.0715
```

```
> fits.med <- median(fitted(model.5))
> fits.groups <- fitted(model.5) <= fits.med
> leveneTest(residuals(model.5),factor(fits.groups))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  8.4328 0.004707 **
```

The tests give very different results in this case, but there is an indication of heteroscedasticity. If this is indeed a problem, the best option to address it is often to use a weighted regression (Kutner et al., 2005, p. 421). This is especially true if the error variance is a function of one of the explanatory variables. A plot of the residuals against each of the explanatory variables (not shown), however, did not reveal any such relationship. Therefore, we will not attempt any weighting.

We saw in Section 13.5 that, based on the output of the function `lm.LMtests()`, the spatial lag model appeared to fit the data better than the spatial error model. We will now fit both of these models and compare them. First, we fit a spatial lag model to the data.

```
> model.5.lag <- lagsarlm(formula.5,
+ data = data.Set4.1, listw = W)
> summary(model.5.lag)
Type: lag
Coefficients: (asymptotic standard errors)
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       5.9514012  1.1511819  5.1698 2.343e-07
Clay             -0.1431834  0.0286813 -4.9922 5.969e-07
SoilP            -0.4732151  0.1295233 -3.6535 0.0002587
I(Clay * SoilP)   0.0124313  0.0033396  3.7224 0.0001974
Weeds            -0.1866198  0.0460086 -4.0562 4.988e-05

Rho: 0.47975, LR test value: 15.146, p-value: 9.9522e-05
Asymptotic standard error: 0.098315
    z-value: 4.8798, p-value: 1.0621e-06
Wald statistic: 23.812, p-value: 1.0621e-06
Log likelihood: -55.11086 for lag model
ML residual variance (sigma squared): 0.19638, (sigma: 0.44314)
Number of observations: 86
Number of parameters estimated: 7
AIC: 124.22, (AIC for lm: 137.37)
LM test for residual autocorrelation
test value: 2.8897, p-value: 0.089145
```

Next, we fit the spatial error model.

```
> model.5.err <- errorsarlm(formula.5,
+    data = data. Set4.1, listw = W)
> summary(model.5.err)
Type: error
Coefficients: (asymptotic standard errors)
                    Estimate Std. Error z value  Pr(>|z|)
(Intercept)        8.8116785  1.1684911  7.5411 4.663e-14
Clay              -0.2153939  0.0291188 -7.3971 1.392e-13
SoilP             -0.5946674  0.1540439 -3.8604 0.0001132
I(Clay * SoilP)    0.0156392  0.0039488  3.9605 7.478e-05
Weeds             -0.2389056  0.0528680 -4.5189 6.216e-06
```

```
Lambda: 0.28077, LR test value: 2.121, p-value: 0.14529
Asymptotic standard error: 0.13267
    z-value: 2.1163, p-value: 0.034318
Wald statistic: 4.4788, p-value: 0.034318
Log likelihood: -61.62322 for error model
ML residual variance (sigma squared): 0.23978, (sigma: 0.48968)
Number of observations: 86
Number of parameters estimated: 7
AIC: 137.25, (AIC for lm: 137.37)
```

The spatial error model and the spatial lag model are not nested, so we cannot compare them directly. The Akaike information criterion is lower for the spatial lag model, so this would lead us to tend to prefer that model, which supports the conclusion reached in Section 13.4 based on the Lagrange Multiplier test. The coefficients of the two models are somewhat different, but not substantially so.

Before moving on, we demonstrate the problem of working with an unscaled *Yield* variable.

```
> data. Set4.1$YieldUS <- yield.pts$Yield
> lagsarlm(YieldUS ~ Clay + SoilP +
+    I(Clay*SoilP) + Weeds, data = data. Set4.1, listw = W)
Error in solve.default(inf, tol = tol.solve):
  system is computationally singular: reciprocal condition number =
1.22513e-13
```

If you get this message with your own data and the simple expedient of scaling does not make it go away, don't despair! There are several alternative things that you can try, which are listed in the Help pages for the functions.

In addition to the spatial error model and the spatial lag model, the `spdep` package also has the capacity to compute a solution to a CAR model (Bivand et al., 2013b, p. 282). This is done using the function `spautolm()`. This function has an argument `family`, which if unspecified, will by default lead to the function fitting a spatial error model. A CAR model solution is obtained by specifying `family = "CAR"`.

```
> W.B <- nb2listw(nlist, style = "B")
> model.5.car <- spautolm(formula.5,
+    data = data. Set4.1, family = "CAR", listw = W.B)
> summary(model.5.car)
Coefficients:
                  Estimate Std. Error z value  Pr(>|z|)
(Intercept)      8.7688570  1.1577547  7.5740 3.619e-14
Clay            -0.2136640  0.0289191 -7.3883 1.488e-13
SoilP           -0.5934295  0.1528116 -3.8834  0.000103
I(Clay * SoilP)  0.0155192  0.0039264  3.9525 7.733e-05
Weeds           -0.2418481  0.0532106 -4.5451 5.491e-06
Lambda: 0.16705 LR test value: 3.0812 p-value: 0.0792
Log likelihood: -61.14311
ML residual variance (sigma squared): 0.22943, (sigma: 0.47899)
Number of observations: 86
Number of parameters estimated: 7
AIC: 136.29
```

Comparing all of the coefficients yields the following.

```
> print(coef(model.5), digits = 2)
    (Intercept)          Clay         SoilP I(Clay * SoilP)         Weeds
          9.319        -0.229        -0.628           0.017        -0.276
> print(coef(model.5.lag), digits = 2)
     rho      (Intercept)    Clay      SoilP   I(Clay * SoilP)       Weeds
0.480          5.951     -0.143     -0.473        0.012            -0.187
> print(coef(model.5.err), digits = 2)
   lambda     (Intercept)    Clay      SoilP   I(Clay * SoilP)       Weeds
0.281         8.812     -0.215     -0.595        0.016            -0.239
> print(coef(model.5.car), digits = 2)
    (Intercept)    Clay      SoilP   I(Clay * SoilP)       Weeds     lambda
      8.769      -0.214     -0.593        0.016            -0.242      0.167
```

In general, the coefficients are similar between the models.

In summary, we have found that the "best" linear regression model as determined by Mallow's $C_p$ indicates spatially autocorrelated errors, and that the indicated error structure is best represented by a spatial lag model of the form (Equation 13.16). It seems counterintuitive that the spatial lag model fits the data better than the spatial error model, since it is unlikely that wheat plants would influence each other at a distance of 61 m. It is very important in this context to remember that correlation does not imply causality. Each of the explanatory variables *SoilP*, *Weeds*, and *Clay* is likely to be autocorrelated. The values of *SoilP* and *Clay* are highly influenced by soil-forming processes, and weed level is likely influenced by the weed seed bank, which likely displays interactive autocorrelation. However, yield response to these autocorrelated explanatory variables, as well as to autocorrelated variables not included in the model, may cause it to display a level of positive autocorrelation sufficient for the lag model to provide the best fit to the data. This is especially true if the uncorrelated errors $\varepsilon$ tend to be larger in magnitude than the correlated errors $\eta$.

### 13.7.2 Comparison of the Mixed Model and Spatial Autoregression

Lambert et al. (2004) compared the generalized least squares (GLS) approach of Section 12.5 and the spatial autoregression (SAR) approach of this section with each other and with OLS, the trend surface (TS) method (Tamura et al., 1988) and the nearest-neighbor (NN) method (Papadakis, 1937), in the analysis of data from a precision agriculture experiment. The latter two methods are discussed in Chapter 16. They found that the GLS and SAR methods were able to detect the interaction of topography with nitrogen in the explanation of yield, while the other three methods were not. The fit of the GLS method was generally considered more accurate than that of the ordinary least squares (OLS) methods but somewhat less accurate than that of the SAR method. Lambert et al. (2004) also found that the standard error of the regression coefficients was considerably less with the GLS and SAR methods than with the other methods. They point out that the GLS method needs more data than the SAR method because it must estimate a variogram. Wall (2004) has criticized the use of the SAR and CAR models for data on an irregular mosaic of polygons. Using an example from the continental United States, she shows that, partly because of the issue of topological invariance discussed in Section 3.5.2 and illustrated in Figure 3.8, these models can lead to counterintuitive results. An alternative for this situation is the mixed model, but unfortunately, the calculation of the experimental variogram necessary to implement this approach is also fraught with difficulties with data on a small, irregular mosaic. As always, the best approach is to try as many methods as possible and attempt to understand the reason behind any difference that appear in the results.

## 13.8 Further Reading

Anselin (1992) provides an excellent, nontechnical introduction to spatial regression. Other good sources are Upton and Fingleton (1985, 1989), Griffith and Layne (1999), Waller and Gotway (2004), and Schnabenberger and Pierce (2001). LeSage and Pace (2009) provide a thorough discussion at a somewhat higher level. Bivand et al. (2013b) contains an extensive discussion and comparison of functions for spatial autoregression. Dormann et al. (2007) presents a comparative study of several different models for dealing with spatially autocorrelated data, and in a supplement several of the authors provide examples of R code to implement these models. A method called the autologistic model has been proposed for the analysis of binary data (Besag, 1974; Augustin et al., 1996). Dormann et al. show that his method has several problems that make it generally less desirable than other methods for the analysis of this type of data.

Lichstein et al. (2002) present a very thorough and extensive comparison of OLS and CAR models, with and without trend terms, for habitat suitability models for southern Appalachian songbirds. Diniz-Filho et al. (2003) and Hawkins et al. (2007) questioned the superiority of spatial regression methods to ordinary least squares in ecological applications. These papers prompted a lively exchange in the literature (Beale et al., 2007; Diniz-Filho et al., 2007). Waller and Gotway (2004) provide a thorough discussion of the issue of heteroscedastic residuals in spatial regression models. Faraway (2002, p. 62) discusses the implementation of weighted regression in R.

For a basic discussion of the Lagrange multiplier test, see Fox (1997) and Theil (1971). Further information is provided by Buse (1982) and Engle (1984).

## Exercises

13.1 Show that the autoregressive time series model $Y_i - \mu = \lambda(Y_i - \mu) + \varepsilon_i$ can be written as a spatial error model.

13.2 The test for spatial autocorrelation in Equation 13.2 was carried out on the model `model.5`, which has the smallest number of explanatory variables of any of the candidate models. Test other models with more explanatory variables for autocorrelated errors. Why might you observe what you do?

13.3 The most complex `leaps()` model developed in Section for Field 4.1 is `model.3`. Test this model to distinguish between the spatial lag model and the spatial error model. Is there a difference between `model.3` and `model.5`?

13.4 Using the spatial lag model, compare the effect of the inclusion of spatial autocorrelation on the coefficients of each of the six candidate models for yield in Field 4.1.

13.5 In this book we always generate autocorrelated data using the method of Haining (1990). However, one can also generate a *Gaussian random field* (Wood and Chan, 1994). The `geoR` package (Ribeiro and Diggle, 2005) contains the function `grf()`, which develops a Gaussian random fields simulation of autocorrelated data. (a) Read about this function and then use it to generate a variable $X$ distributed as a Gaussian random field on a 20 by 20 grid; (b) Use the `geoR` function `variog()` to display the experimental variogram of $X$.

13.6 Use the Gaussian random field data from Exercise 13.4 to generate a binomially distributed random variable $Y$ based on a logistic probability on $X$. Then use the GLM model to fit these data and compute the bias in the fit.