3. The Spatial Autoregressive Model

Given the above formulation of spatial structure in terms of weights matrices, our objective in this section is to develop the basic model of areal-unit dependencies that will be used to capture possible spatial correlations between such units. Unless otherwise stated, we shall implicitly represent the relevant set of areal units, $\{R_1,..,R_n\}$, by their indices, $i = 1,..,n$. In particular, these areal units will almost always represent the *sample units* of interest. To put this spatial-dependency model in proper perspective, we begin with a typical linear model of the form

$$(3.1) \qquad Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i \ , \ i = 1,..,n$$

where Y_i is taken to represent some relevant attribute of each spatial unit, i, and where $(x_{ij} : j = 1,..,k)$ represents a set of "explanatory" attributes of i that are postulated to influence Y_i. For example, if Y_i is the Myocardial Infarction rate of each English Health District, $i = 1,..,190$, in Section 1.3 above, then x_{i1} might correspond to the Jarman score for District i, together with other possible attributes of that district. This model exhibits an obvious similarity to expression (7.5) in Part II. The key difference is in terms of their respective *spatial sample units*, where the point locations (s) in expression (7.5) are here replaced by areal units (R) that partition this space. As mentioned in the introduction, this change in spatial sample units reflects the type of spatial data being analyzed. For example, while, say, temperature is meaningful each point in space, this is not true of Myocardial Infarction rates.[1] But much more important for our present purposes is the way in which the *unobserved errors* (or *residuals*) are treated in each model. Notice in particular that we have switched notation in (3.1), and are now representing such residuals by u_i rather than ε_i. The reason for this is that we shall proceed to develop an explicit linear model of these spatial residuals themselves.

Before doing so, it is convenient to restate (3.1) in matrix terms as

$$(3.2) \qquad Y = X\beta + u$$

where as usual, $Y = (Y_1,..,Y_n)', X = [1_n, x_1,..,x_k], \beta = (\beta_0, \beta_1,.., \beta_k)$ and $u = (u_1,..,u_n)'$. We again assume that the random vector, u, of residuals is multinormally distributed with mean, $E(u) = 0$, so that by construction,

$$(3.3) \qquad E(Y) = X\beta$$

[1] Note however that in cases such as the California rainfall example, where cities were treated as points, the relevant data implicitly involves "local" spatial averages. So in this setting, for example, it would be perfectly meaningful to compare the Myocardial Infarction rates of San Francisco and Los Angeles.

In this setting, our primary objective is to model the covariance structure of u in a manner that reflects possible spatial dependencies among areal units.

But rather than postulate spatial stationarity properties of u (as was done for spatially continuous data in Part II), we must now rely on discrete spatial structure as summarized by a given spatial weights matrix, $W = (w_{ij} : i, j = 1,..,n)$. In terms of our Myocardial example above, w_{ij}, may represent some measure of the spatial proximity of Health District j to (or influence on) Health District i, where higher values of w_{ij} denote greater spatial proximity or influence. In this setting, it seems reasonable to postulate that each unobserved residual, u_i, in (3.1) is influenced by those residuals, u_j, in neighboring areal units j, i.e., with positive spatial weights, w_{ij}. As a parallel to (3.1), such influences might also be represented by linear "spatial error" model of the form:

$$(3.4) \qquad u_i = \sum_{j \neq i} \alpha(w_{ij}) u_j + \varepsilon_i$$

where $\alpha(w_{ij})$ is some appropriate "influence" function depending on w_{ij}, and where ε_i represents that part of residual u_i that is not influenced by other areal units. But as we have seen in Section 3.2, there is already great flexibility in the specification of spatial weights, w_{ij}, and hence no need for further functional elaborations. Rather, the strategy here is to use the simplest possible specification in terms of a common scale factor, ρ, so that $\alpha(w_{ij})$ takes the form ρw_{ij}, and (3.4) reduces to[2]

$$(3.5) \qquad \boxed{u_i = \rho \sum_{j \neq i} w_{ij} u_j + \varepsilon_i \quad , \quad i = 1,..,n}$$

To interpret (3.5), note first that (except for the absence of an intercept term) this relation is essentially a type of linear regression model in which each residual, u_i, is regressed on its neighbors, u_j (with coefficients ρw_{ij}). Moreover, since this effectively implies that the full set of residuals is being regressed on itself, model (3.5) is designated as a *spatial autoregressive model* of residual dependencies. In this context, the summation over all $j \neq i$ ensures that no individual residual is "regressed on itself". But even with this restriction, it will be shown below that the estimation of such autoregressive models is far more subtle than that of standard regression models.

For the present however, we focus only on the basic meaning of (3.5). First consider the parameter, ρ, which plays a very special role in this model. At one extreme, if $\rho = 0$ then each residual, u_i, reduces to its own *intrinsic component*, ε_i, and all spatial dependencies vanish. More formally, if we now assume that these individual components are independently and identically normally distributed as,

[2] Here the notation, $\Sigma_{j \neq i}$, means summation over all units, j, other than unit i.

(3.6) $$\varepsilon_i \sim N(0,\sigma^2)\ ,\ \ i=1,..,n$$

then model (3.1) is seen to reduce to a standard linear regression model when $\rho = 0$. At the other extreme, when $|\rho|$ becomes large, the strength of all spatial dependencies (positive or negative) must also become large. This suggests that ρ be designated as the *spatial dependency parameter* for the model.

Note also, that for any pairs of areal units, ij and kh, with positive spatial weights, $w_{ij}, w_{kh} > 0$, and any nonzero level of spatial dependence, $\rho \neq 0$, it must always be true that

(3.7) $$\frac{\rho w_{ij}}{\rho w_{kh}} = \frac{w_{ij}}{w_{ik}}$$

Thus the relative strength of these spatial dependencies is determined entirely by their spatial weights. In summary, this model provides a natural "division of responsibilities" in which ρ governs the *overall strength* of spatial dependencies, and in which the spatial weight structure governs their *relative strength* among individual areal-unit pairs.

Finally, to write this model in more compact matrix form, it convenient to assume that $w_{ii} = 0$ in the given spatial weights matrix, W, so that (3.5) can be rewritten in more standard terms as

(3.8) $$u_i = \rho \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i \ \ ,\ \ i=1,..,n$$

In this form, if we now let $\varepsilon = (\varepsilon_1,..,\varepsilon_n)'$ denote the random vector of intrinsic components, then expressions (3.8) and (3.6) together yield the follows *Spatial Autoregressive Model* of residual dependencies:[3]

(3.9) $$u = \rho W u + \varepsilon \quad , \quad \varepsilon \sim N(0,\sigma^2 I_n)$$

where in addition it is assumed that the diagonal element of W are zero, written as

(3.10) $$diag(W) = 0\ .$$

3.1 Relation to Time Series Analysis

Like most of the spatial dependency models considered in these notes, model (3.9) was originally inspired by a time series model [as in Whittle (1954)]. In the present case, this

[3] This model was originally proposed by Whittle (1954). But the present matrix formulation was first given by Ord (1975), who designated (3.9) as a *first-order* spatial autoregressive process.

"parent" model can be formulated as follows. If we consider a finite sequence of random variables, $(u_t : t = 1,..,T)$, over T time periods (say average Philadelphia temperature. u_t, over T successive days), then the standard *first-order autoregressive* [AR(1)] *model* of this series takes the recursive form:

(3.1.1) $$u_t = \rho u_{t-1} + \varepsilon_t \ , \ t = 2,..,T$$

with "initial condition",[4]

(3.1.2) $$u_1 = \varepsilon_1$$

where $(\varepsilon_t : t = 1,..,T)$ is assumed to be a sequence of independent random "innovations" identically distributed as $N(\mu, \sigma^2)$. In the "temperature" example above, these innovations $(\varepsilon_t : t = 1,..,T)$ can be viewed as random fluctuations about some constant mean daily temperature, μ. The term "first-order" in this case refers to the fact that given the past history of daily temperatures in Philadelphia, model (3.1.1) assumes that today's temperature, u_t, depends only on yesterday's temperature, u_{t-1} plus some current temperature innovation, ε_t.

Except for the nonzero value of μ, this AR(1) model can be viewed formally as a special case of model (3.9). To see this, observe simply that if the $T \times T$ weights matrix, $W = (w_{ts} : t, s = 1,..,T)$, is defined by

(3.1.3) $$w_{ts} = \begin{cases} 1 & , t = 2,..,T,\ s = t-1 \\ 0 & , otherwise \end{cases}$$

then it follows at once from (3.1.1) and (3.1.2) that:

(3.1.4) $$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} = \rho \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix} \Rightarrow u = \rho W u + \varepsilon$$

But this particular instance of (3.9) has the important property that time dependencies flow in only *one direction* – namely from the past to the present. Formally, this is reflected by the so-called "lower triangular" structure of W in (3.1.4).

[4] While (3.1.2) can be replaced by more standard "steady state' initial conditions, the present simpler form is most appropriate for our purposes.

To appreciate the significance of this unidirectional flow, it is instructive to ask how one might *simulate* this model. Here the answer is almost self-evident from (3.1.1) and (3.1.2):

Step 1: Sample a value of ε_1 from $N(\mu,\sigma^2)$ and set $u_1 = \varepsilon_1$.

Step 2: Sample a value of ε_2 from $N(\mu,\sigma^2)$ and set $u_2 = \rho u_1 + \varepsilon_2$.

Step 3: Sample a value of ε_3 from $N(\mu,\sigma^2)$ and set $u_3 = \rho u_2 + \varepsilon_3$.

•
•
•

Step *T*: Sample a value of ε_T from $N(\mu,\sigma^2)$ and set $u_T = \rho u_{T-1} + \varepsilon_T$.

However, for more general examples of model (3.9), this simple process of simulation is not possible.

3.2 The Simultaneity Property of Spatial Dependencies

This problem is mostly easily illustrated by the following one-dimensional example. Suppose we consider "over the fence" communications between residential neighbors on a given street, as depicted in Figure 3.1 below.

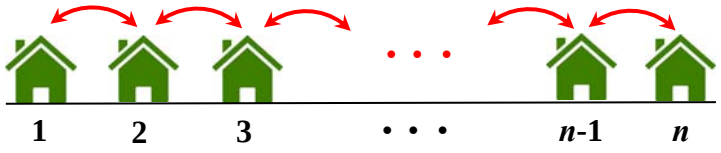


Figure 3.1. Bilateral Dependency Example

In particular, suppose that household i's opinion, u_i, on how much each house should contribute to their annual street party is influenced both by i's initial opinion, ε_i, and by the opinions of i's immediate neighbors, including u_{i-1} and/or u_{i+1}. Then a natural spatial model of opinion formation by these residents might well take the form:

(3.2.1)
$$u_i = \begin{cases} \rho u_{i+1} + \varepsilon_i & , \ i = 1 \\ \rho(u_{i-1} + u_{i+1}) + \varepsilon_i & , \ 2 \le i \le n-1 \\ \rho u_{i-1} + \varepsilon_i & , \ i = n \end{cases}$$

where ρ now reflects how influential the opinions of these neighbors are. Note in particular that the "edge" residents 1 and n have only one neighbor, while all other residents have two neighbors.

Given this spatial model of opinion formation,[5] one may again ask: how might we simulate this model? Here the key question is where to *start* the simulation. For if we start with edge resident 1, then it clear from the first line of (3.2.1) that we must know the opinion, u_2, of 1's neighbor in order to simulate u_1. Similarly, if we start with edge resident n then the last line of (3.2.2) shows that the opinion, u_{n-1}, of n's neighbor is required to simulate u_n. Moreover, the situation is even worse for intermediate residents, i, where both neighboring opinions, u_{i-1} and u_{i+1}, are required in order to simulate u_i. So it would appear that there is no way to simulate this process at all. But to be more precise, this argument shows that there is no possible *sequential simulation* procedure for realizing samples of (3.2.1). Rather, the full set of opinions, $(u_1,u_2,..,u_n)$, must be somehow be simulated *simultaneously*.

Here it turns out that there is a remarkably simple procedure for doing so. In particular, let us again formulate (3.2.1) as an instance of (3.9) where W now takes the form:

$$(3.2.2) \qquad w_{ts} = \begin{cases} 1 & , t = 2,..,n-1,\ s \in \{t-1, t+1\} \\ 0 & , otherwise \end{cases}$$

then it follows at once from (3.1.1) and (3.1.2) that:

$$(3.2.3) \qquad \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} = \rho \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{n-1} \\ \varepsilon_n \end{pmatrix} \Rightarrow u = \rho W u + \varepsilon$$

But given this matrix formulation, observe that we may solve for u in terms of ε as follows:

$$(3.2.4) \qquad u = \rho W u + \varepsilon \ \Rightarrow \ u - \rho W u = \varepsilon$$

$$\Rightarrow \ (I_n - \rho W)u = \varepsilon$$

So assuming for the moment that the inverse matrix, $(I_n - \rho W)^{-1}$, *exists*, we can multiply both sides of (3.2.4) by $(I_n - \rho W)^{-1}$ to obtain the following *reduced form solution* for u in terms of ε,

$$(3.2.5) \qquad \boxed{u = (I_n - \rho W)^{-1}\varepsilon}$$

[5] Formally, expression (3.2.1) is an instance of the *bilateral autoregressive process* proposed by Whittle (1954). Indeed, this is precisely the one-dimensional example that motivated his original analysis of spatial autoregressive processes.

Given this existence assumption, observe that if "intrinsic opinions" are again assumed (for sake of illustration) to be independently and identically normally distributed about some average opinion level, μ, as $\varepsilon_i \sim N(\mu,\sigma^2)$, $i=1,..,n$, then we can now simulate (3.1.5) in essentially only two steps:

Step 1: Sample each ε_i from $N(\mu,\sigma^2)$, $i=1,..,n$, and set $\varepsilon=(\varepsilon_1,..,\varepsilon_n)'$.

Step 2: Solve for $u=(u_1,..,u_n)'$ as $u=(I_n-\rho W)^{-1}\varepsilon$.

So by simple matrix manipulations, this simultaneity problem appears to have been solved. But there remains the question of how this "magic" was possible, and what it actually means in more intuitive terms.

3.3 A Spatial Interpretation of Autoregressive Residuals

Our objective in this section is to obtain conditions for the existence of $(I_n-\rho W)^{-1}$ and to give an intuitive spatial interpretation to this inverse matrix. To do so, we start by recalling that for any number, *a*, the basic *geometric series*:

(3.3.1) $$S=1+a+a^2+a^3+\cdots=\sum_{k=0}^{\infty}a^k$$

represents the simplest example of an infinite summation that can be given a closed form solution in an elementary way. For if one considers the partial sum,

(3.3.2) $$S_k=1+a+a^2+a^3+\cdots+a^k$$

and multiplies this by *a*,

(3.3.3) $$aS_k=a+a^2+a^3+\cdots+a^k+a^{k+1}$$

then by subtracting (3.3.3) from (3.3.2),

(3.3.4) $$S_k-aS_k=(1+a+a^2+a^3+\cdots+a^k)-(a+a^2+a^3+\cdots+a^k+a^{k+1})=1-a^{k+1}$$

we obtain the simple identity

(3.3.4) $$S_k=\frac{1-a^{k+1}}{1-a}$$

But since by definition, $S=\lim_{k\to\infty}S_k$, it follows at once from (3.3.4) that this limiting sum exists if and only if $\lim_{k\to\infty}a^k=0$, and must have the closed-form solution:

(3.3.5) $$S = \lim_{k\to\infty} S_k = \frac{1}{1-a} = (1-a)^{-1}$$

Finally, by combining (3.3.1) and (3.3.5) we see that

(3.3.6) $$(1-a)^{-1} = 1 + a + a^2 + a^3 + \cdots = \sum_{k=0}^{\infty} a^k$$

if and only if $\lim_{k\to\infty} a^k = 0$.

The point of this exercise for our purposes is that exactly the same argument can be applied to matrices, by simply substituting the scalar, *a*, with an *n*-square matrix *A*. In particular, if O_n denotes the n-square zero matrix, then it is shown in Section A3.5.2 of the Appendix that

(3.3.7) $$(I - A)^{-1} = I_n + A + A^2 + A^3 + \cdots = \sum_{k=0}^{\infty} A^k$$

if and only $\lim_{k\to\infty} A^k = O_n$. So in our case, by setting $A = \rho W$, it follows that that the inverse $(I_n - \rho W)^{-1}$ will exist and have the limiting form

(3.3.8) $$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \cdots = \sum_{k=0}^{\infty} \rho^k W^k$$

if and only if

(3.3.9) $$\lim_{k\to\infty} \rho^k W^k = O_n$$

Our main objective is to employ this representation to give a meaningful interpretation to the "steady states" of spatial autoregressive processes as in expression (3.2.5). But before doing so, it is important to establish conditions on the spatial dependency parameter which will ensure that (3.3.9) holds. Since this condition must surely hold when $\rho = 0$, it is not surprising that the desired condition will amount to placing a bound on the maximum size of $|\rho|$. But this bound will of course depend on the structure of the spatial weights matrix, *W*, as we now show.

3.3.1 Eigenvalues and Eigenvectors of Spatial Weights Matrices

In Section A3.1 of the Appendix we develop a number of important properties of *n*-square matrices, *A*, as representations of *n*-dimensional linear transformations on $\mathbb{R}^n$. Our focus is on the geometric interpretations of these properties, which can often be represented graphically in 2 dimensions. Without going into great detail here, it is enough to say that every 2-square matrix,

(3.3.10) $$A = (a_1, a_2) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

represents a 2-dimensional linear transformation that transforms each vector, $x = (x_1, x_2) \in \mathbb{R}^2$, into to a new vector, $Ax \in \mathbb{R}^2$, called the *image* of x under A. Each transformation, A, is entirely representable by the images of the *identity basis vectors*, $e_1, e_2 \in \mathbb{R}^2$ [recall expression (3.2.16) if Part II], as shown in Figure 3.2. In particular, since by definition each $x = (x_1, x_2)$ is representable as the weighted sum, $x = x_1 e_1 + x_2 e_2$, it follows from linearity that Ax is representable by the corresponding weighted sum of the images, (Ae_1, Ae_2), as shown in Figures 3.3 below (see also Figures A3.3 and A3.4 in the Appendix).

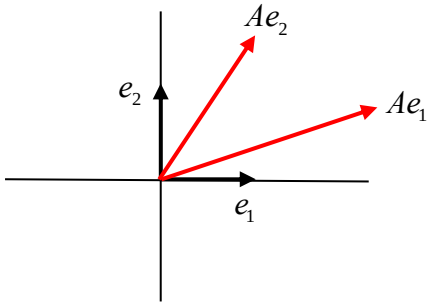


Figure 3.2. Basis Image Vectors

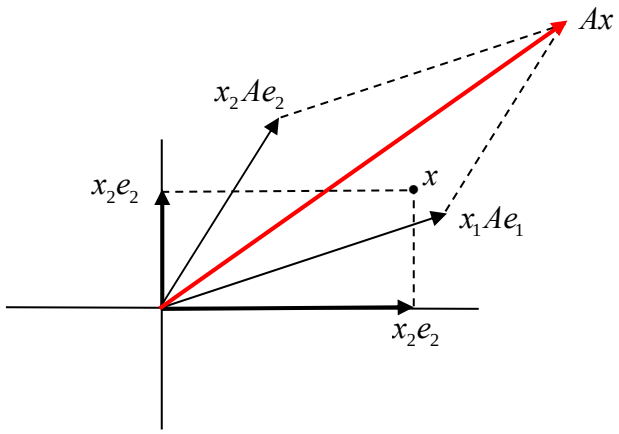


Figure 3.3. General Image Vectors

From a geometrical viewpoint, it is of interest to ask whether there exist any vectors, $x \in \mathbb{R}^n$, that are simply "stretched" by A into (possibly negative) multiples of themselves, i.e., whether

(3.3.11) $$\boxed{Ax = \lambda x}$$

for some scalar, $\lambda \in \mathbb{R}$. If so, then λ is called an *eigenvalue* of A with associated *eigenvector*, x. [Note that (3.3.11) continues to hold for any scalar multiple of x, so that eigenvectors are only unique up to scalar multiples.] For convenience we refer to eigenvalues together with their eigenvectors as the *eigenstructure* of A, and in particular, denote the set of distinct eigenvalues for A by $Eig(A)$. To illustrate these ideas for *spatial weights* matrices in 2 dimensions, we are of course restricted to the simplest possible case of only two areal units, as shown in Figure 3.4 below.

(3.3.12)

R_1	R_2

$$W = \begin{pmatrix} 0 & w_{12} \\ w_{21} & 0 \end{pmatrix}$$

If W represents a simple contiguity relation with $w_{12} = 1 = w_{21}$ [as in the 3-unit example of expression (2.1.22) above], and if we let $x_1 = (1, 1)'$ and $x_2 = (-1, 1)'$, then simple matrix multiplication shows that $W x_1 = x_1$ and $W x_2 = -x_2$, so that these are both eigenvectors of W with corresponding eigenvalues, $Eig(W) = \{\lambda_1, \lambda_2\} = \{1, -1\}$. This is shown graphically in Figure 3.4 below (where x_1 and Wx_1 are slightly offset so that both can be seen):

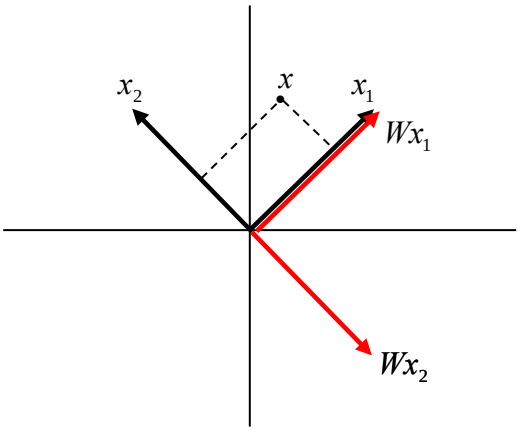


Figure 3.4. Eigenstructure of W

More generally (as shown in Section A3.3 of the Appendix), each *n*-square matrix, *A*, possesses at most *n* distinct eigenvalues. To see that there may be fewer than *n*, consider the identity matrix, I_n, which has only one distinct eigenvalue $(\lambda = 1)$ since by definition, $I_n x = x$ for all $x \in \mathbb{R}^n$. This example also shows that eigenvectors in such cases can be chosen in many ways. There also exist matrices with *no* (real) eigenvalues, as illustrated by the matrix

(3.3.13) $$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

As seen in Figure 3.5 below, this matrix rotates the plane by $90°$, so that no vector can be sent into a scalar multiple of itself.

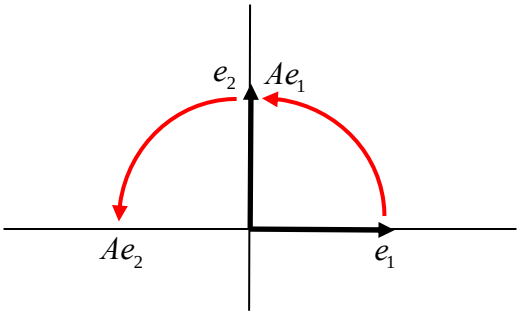


Figure 3.5. Rotation Transformation

But for sake of simply, we focus here *n*-square matrices, A, with a *full set* of eigenvalues, $Eig(A) = \{\lambda_1,..,\lambda_n\}$, and associated eigenvectors, $x_1,..,x_n$, that are *linearly independent*.[6] In geometric terms, this means that every point, $x \in \mathbb{R}^n$, can be written as a linear combination of these eigenvectors, as illustrated by the point, x, in Figure 3.4. In algebraic terms, it means that the *n*-square matrix, $X = [x_1,..,x_n]$, defined by these eigenvectors is nonsingular, so that the inverse matrix, X^{-1}, exists. We may thus write out the relations among these eigenvalues and eigenvectors as follows,

(3.3.14) $$A x_i = \lambda_i x_i \ , \ i = 1,..,n$$

$$\Rightarrow A X = [Ax_1,..,Ax_n] = [\lambda_1 x_1,..,\lambda_n x_n] = [x_1,..,x_n]\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$$\Rightarrow \boxed{A X = X \Lambda}$$

where $\Lambda = diag(\lambda_1,..,\lambda_n)$ is the diagonal matrix of eigenvalues. So (post) multiplying both sides of (3.3.14) by X^{-1}, we obtain the following "spectral" representation of A,

(3.3.15) $$A X X^{-1} = X \Lambda X^{-1} \Rightarrow \boxed{A = X \Lambda X^{-1}}$$

To see the power of this representation, observe that if we multiply A by itself, then:

(3.3.16) $$A^2 = (X \Lambda X^{-1})(X \Lambda X^{-1}) = X \Lambda (X^{-1} X) \Lambda X^{-1} = X \Lambda^2 X^{-1}$$

By comparing this with (3.3.15), it follows at once that that the eigenvalues of A^2 are precisely the *squares* of the eigenvalues of A, and moreover that the associated eigenvectors remain the *same*. By simply repeating this argument k times, it follows more generally that

(3.3.17) $$A^k = X \Lambda^k X^{-1} = X \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} X^{-1} \ , \quad k = 1,2,...$$

So the eigenstructure of A tells us a great deal about how the associated powers, A^k, of A must behave. In particular, the *limiting behavior* of these powers as $k \to \infty$ for *any* matrix, A, is governed entirely by the *maximum size* of its eigenvalues, which we denote by,[7]

(3.3.18) $$|\lambda|_A = \max_{\lambda \in Eig(A)} |\lambda| ,$$

[6] In fact the eigenvectors for distinct eigenvalues are *always linearly independent*, as illustrated in Figure A3.27 of the Appendix.

[7] As discussed in Section A3.5 of the Appendix, this maximum absolute value is usually referred to as the *spectral radius* of the matrix, A.

To see this, note simply from (3.3.17) that these powers will converge to the zero matrix if and only if $\lambda^k \to 0$ for *all* $\lambda \in Eig(A)$. Because this is equivalent to the single condition, $|\lambda|_A < 1$, it then follows that

(3.3.18) $$\boxed{\lim_{k\to\infty} A^k = O_n \;\Leftrightarrow\; |\lambda|_A < 1}$$

For the important case of *nonnegative* matrices, it shown in Section A3.5.1 of the Appendix that this maximum always corresponds to the largest *positive* eigenvalue of A, denoted here by λ_A, so that $\lambda_A = |\lambda|_A$. As an illustrative example, the eigenstructure of the nonnegative matrix,

(3.3.19) $$A = \begin{pmatrix} 2/3 & 1/3 \\ 1/6 & 1/2 \end{pmatrix}$$

is easily seen to be given by

(3.3.20) $$\Lambda = \begin{pmatrix} 5/6 & \\ & 1/3 \end{pmatrix} \quad , \quad X = [x_1, x_2] = \left[\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right]$$

(as can be checked by matrix multiplication). This eigenstructure is shown graphically in Figure 3.6 below.

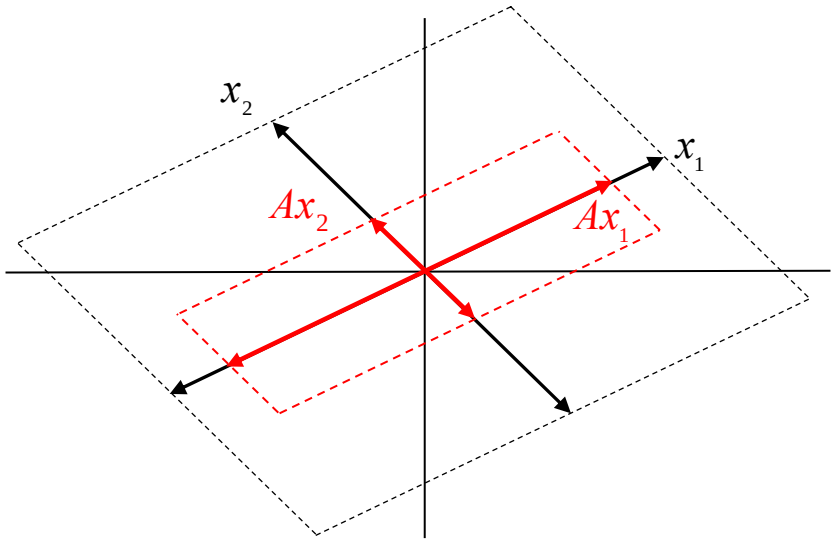


Figure 3.6. "Shrinking" Eigenvalue Example

Since all points are linear combinations of the eigenvectors, x_1 and x_2, and since $|\lambda|_A = \lambda_A = 5/6 < 1$ implies that both these eigenvectors shrink toward zero, we see that

all points are shrunk towards zero (as illustrated by the parallelogram in the figure). In other words, by using the coordinate system created by these eigenvectors, we see that the shrinking behavior of these eigenvectors is inherited by all points *with respect to this coordinate system*. While not every case is so simply illustrated, Figure 3.6 helps to provide some geometric intuition for the general result in (3.3.18).[8]

3.3.2 Convergence Conditions in Terms of ρ

By combining (3.3.9) and (3.3.18), we see that a necessary and sufficient condition for the geometric-series representation in (3.3.8) to hold is that the maximum eigenvalue of the matrix (ρW), be less than one. But for each eigenvalue, λ, of W, say with eigenvector, x, it follows at once from (3.3.11) that

(3.3.21) $$Wx = \lambda x \;\Rightarrow\; \rho Wx = \rho\lambda x \;\Rightarrow\; (\rho W)x = (\rho\lambda)x$$

and thus that $\rho\lambda$ is automatically an eigenvalue for (ρW), so that

(3.3.22) $$Eig(\rho W) = \rho\, Eig(W)$$

In particular, since this implies that

(3.3.23) $$|\lambda|_{\rho W} = |\rho||\lambda|_W = |\rho|\,\lambda_W$$

it follows that

(3.3.24) $$|\lambda|_{\rho W} < 1 \;\Leftrightarrow\; |\rho|\,\lambda_W < 1 \;\Leftrightarrow\; |\rho| < \frac{1}{\lambda_W}$$

So for the present case of spatial weight matrices, W, the general convergence condition in (3.3.18) now takes the form

(3.3.25) $$\lim_{k\to\infty} \rho^k W^k = O_n \;\Leftrightarrow\; |\rho| < 1/\lambda_W$$

so that by (3.3.8) and (3.3.9),

(3.3.26) $$\boxed{(I_n - \rho W)^{-1} = \sum_{k=0}^{\infty} \rho^k W^k \;\Leftrightarrow\; |\rho| < 1/\lambda_W}$$

Note in particular that if the maximum eigenvalue of W happens to be unity, i.e., $\lambda_w = 1$, then (3.3.25) takes the simple and appealing form[9]

[8] See Section A3.5.3 in the Appendix for a general development of this result.

[9] Here it must be stressed that in spite of the apparent similarity of the condition, $|\rho| < 1$, to the properties of correlation coefficients, this spatial dependency parameter, ρ, is *not* a correlation coefficient.

(3.3.27) $$\boxed{(I_n - \rho W)^{-1} = \sum_{k=0}^{\infty} \rho^k W^k \Leftrightarrow |\rho| < 1}$$

For this reason, it is often convenient to normalize W to have a maximum eigenvalue of one. The simplest procedure for doing so is to divide W by it maximum eigenvalue, λ_W, say $W^* = \frac{1}{\lambda_W} W$. For this normalized weights matrix, it then follows from the same argument in (3.3.21) through (3.3.23) that

(3.3.28) $$Eig(W^*) = Eig\left(\tfrac{1}{\lambda_W} W\right) = \tfrac{1}{\lambda_W} Eig(W) \;\Rightarrow\; \lambda_{W^*} = \tfrac{1}{\lambda_W}(\lambda_W) = 1$$

and thus that (3.3.27) always holds for $W = W^*$. In fact, this is the primary motivation for the normalizing convention in expression (2.1.25) of Section 2 above.

Before proceeding, it is important to note that *row normalized* weight matrices, W_{rn}, must also exhibit this same property. This can be seen in part by observing that the normalizing condition (2.1.19) for W_{rn} in Section 2 can be written as

(3.3.29) $$1 = \sum_{j=1}^{n} w_{ij} = [w_{i1},..,w_{in}] \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = w_i' 1_n \;\; , \;\; i = 1,..,n$$

where w_i' is the i^{th} row of W_{rn}. This set of conditions can in turn be written in matrix form as

(3.3.30) $$\begin{pmatrix} w_1' \\ \vdots \\ w_n' \end{pmatrix} 1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1_n \;\Rightarrow\; \boxed{W_{rn} 1_n = 1_n},$$

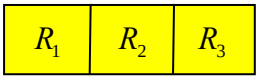
which shows that 1_n must always be an eigenvector of W_{rn} with *unit eigenvalue*. Thus for the row normalization of any spatial weights matrix, W, we must have $1 \in Eig(W_{rn})$. In addition, it is shown in Section A3.5.2 of the Appendix that this unit eigenvalue is necessarily the maximum eigenvalue of W_{rn}, and thus that (3.3.27) must always hold for row normalized matrices.

3.3.3 A Steady-State Interpretation of Spatial Autoregressive Residuals

Assuming that ρW satisfies (3.3.25), it remains to give a spatial interpretation of the expanded representation of $(I_n - \rho W)^{-1}$ in (3.3.8). To do so, it is useful to start by considering the *direct influences* among areal units as implied by a given spatial weights

matrix, W. This is well illustrated by the example in expression (2.1.22) of Section 2, which we reproduce here for convenience,

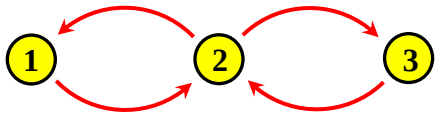
(3.3.31)



$$W = \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

In this example, the only direct influences are between unit 2 and each of the other units, 1 and 3. This can be represented by the following graph, with areal units as "nodes" and positive weights as directed "links" (in red):

(3.3.32)



So, for example, the top two arrows show that unit 2 directly influences both units 1 and 3. Now consider the square of this weight matrix,

(3.3.33) $$W^2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

If one thinks of direct links as *influence paths* of *length* 1, then the ij elements of $W^2 = (w_{ij}^2)$ are precisely number of influence paths of *length* 2 from j to i. In particular, each m^{th} term of the ij-value, $w_{ij}^2 = \Sigma_{m=1}^3 w_{im} w_{mj}$, of W^2 contributes a value of 1 to this sum if and only if both w_{im} and w_{mj} are 1, i.e., if and only if there is a path, $j \to m \to i$, of length 2. For example, while unit 3 does not directly influence unit 1, there is an *indirect influence* on the path, $3 \to 2 \to 1$, seen in (3.3.32). This single influence path of length 2 corresponds to the 1 in the upper right hand corner of W^2. Notice also that while the diagonal elements of W are zero by construction, this is not true of W^2. For example there is now an influence path of length 2 from unit 1 to itself, namely the path $1 \to 2 \to 1$ in which 1's influence on 2 is "echoed back" as a second order influence on 1. In a similar manner, the ij elements of the k^{th} power, $W^k = (w_{ij}^k)$, of W indicate the number of length k paths from j to i. But notice in the present example, that these relations depend explicitly on the fact that W consists entirely of zeroes and ones. More generally, for any n-square weights matrix, W, the ij elements of the k^{th} power, $W^k = (w_{ij}^k)$, of W take the form[10]

(3.3.34) $$w_{ij}^k = \Sigma_{m_1=1}^n [\Sigma_{m_2=1}^n [\cdots [\Sigma_{m_{k-1}=1}^n w_{im_1} w_{m_1 m_2} \cdots w_{m_{k-1} j}] \cdots]]$$

[10] For a deeper discussion of such influence paths see Martellosio (2012).

where each positive product, $w_{im_1}w_{m_1m_2}\cdots w_{m_{k-1}j}$, in w_{ij}^k still corresponds to a unique path, $j \to m_{k-1} \to \cdots \to m_1 \to i$, of positive influences – but where this product need not be unity. Moreover, if we now introduce the *spatial dependency parameter*, ρ, and consider the k^{th} power, $\rho^k W^k$, then (3.3.34) becomes

$$(3.3.35) \qquad \rho^k w_{ij}^k = \Sigma_{m_1=1}^n[\Sigma_{m_2=1}^n[\cdots[\Sigma_{m_{k-1}=1}^n(\rho w_{im_1})(\rho w_{m_1m_2})\cdots(\rho w_{m_{k-1}j})]\cdots]]$$

In this form, it is clear that the w-values along each path reflect only the *relative* influences of each link, where typically such influences will be smaller on links between more widely separated units. The *full* influences of these links are then determined by ρ.

With these preliminary observations, it should now be clear that the geometric sum in (3.3.8) represents the cumulative effect of all these direct and indirect spatial influences among units. The can be seen more explicitly by using (3.3.8) to expand (3.2.5) as follows:

$$(3.3.36) \qquad \boxed{\begin{aligned} u = (I_n - \rho W)^{-1}\varepsilon &= (I_n + \rho W + \rho^2 W^2 + \cdots)\varepsilon \\ &= \varepsilon + \rho W\varepsilon + \rho^2 W^2\varepsilon + \cdots \end{aligned}}$$

So for any given vector of intrinsic effects, $\varepsilon = (\varepsilon_1,..,\varepsilon_n)'$, expression (3.3.36) displays the accumulation of all direct and indirect effects of ε that define the vector, $u = (u_1,..,u_n)'$, of autoregressive residuals. This is illustrated graphically in Figure 3.7 below for the "over the fence" communications example in Figure 3.1 (for the case of $n = 7$ neighbors):

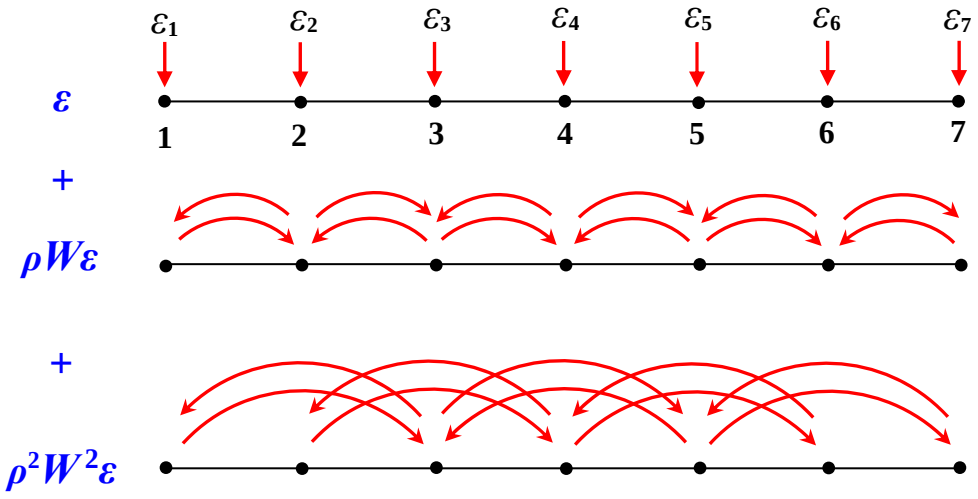


Figure 3.7. Spatial Ripple Effect

Here we only show the first three terms of (3.3.35), where the first term reflects the initial (intrinsic) opinions of each neighbor, and where subsequent terms represent the cumulative indirect influences on these opinions resulting from over-the-fence communications. Alternatively, if one were to imagine each initial opinion as a pebble falling into water, then the influences of these opinions spread out like "ripples" in all directions. (An empirical example of such a *ripple effect* is given in Figure 7.8 below.)

More generally this example suggests that spatial autoregressive residuals, u, can be viewed as the *steady state* of an implicit spatial diffusion process generated by a random vector of intrinsic effects, ε. Of course, the spatial autoregressive model in (3.9) is *static* in nature, and involves no explicit notion of time. But such cumulative effects can nonetheless be usefully represented as a steady state over virtual time periods as shown in Figure 3.8 below.

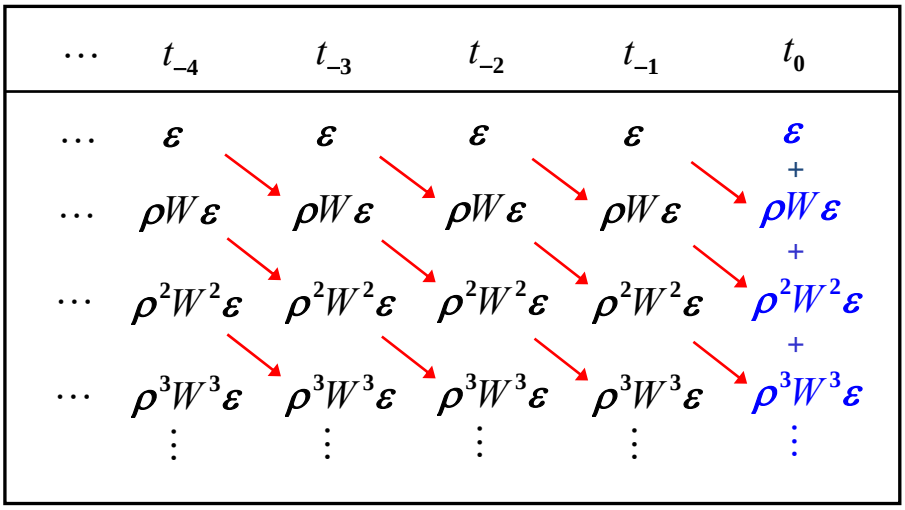


Figure 3.8. Steady State Interpretation

Here for example, $\rho W\varepsilon$, in the "current" state, t_0, is interpreted as the direct effect of ε in the "previous" state, t_{-1}, and similarly, $\rho^2W^2\varepsilon$, in t_0 is the indirect effect of ε in t_{-2}. The main feature of this representation is that the total effect in each state resulting from all previous states remains the same, thus yielding a "steady state" independent of time.

But regardless of whether or not this steady state interpretation is used, the essential result here is that the reduced-form representation of spatial autoregressive residuals, *u*, in (3.3.35) does indeed incorporate all direct and indirect effects generated by ε in the presence of spatial structure, ρW.

One final point needs to be made about this reduced-form representation. It is often observed that this representation is not essential for the existence of the inverse $(I_n - \rho W)^{-1}$. For example, if W is given by (3.3.30), and say, $\rho = 2$, then it may be

verified (by direct matrix multiplication) that the inverse of this matrix exists, and is given (approximately) by

$$(3.3.37) \qquad (I_3 - \rho W)^{-1} = \begin{pmatrix} 0.4286 & -0.2857 & -0.5714 \\ -0.2857 & 0.1429 & -0.2857 \\ -0.5714 & -0.2857 & 0.4286 \end{pmatrix}$$

But while this inverse exists, it is far more difficult to interpret in a meaningful way. In particular, the negative elements in this matrix are rather questionable. Note in particular from the positivity of ρ that ρW must be a nonnegative matrix. So it seems clear from the basic autoregressive relation, $u = \rho W u + \varepsilon$, that a positive increase in the components of ε in the should certainly not decrease any component of u. However, (3.3.37) and (3.3.36) together imply for example that the second component, u_2, is related linearly to $\varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3)'$ by

$$(3.3.38) \qquad u_2 = -(0.2857)\,\varepsilon_1 + (0.1429)\,\varepsilon_2 - (0.2857)\,\varepsilon_3$$

Thus an increase in either ε_1 or ε_3 will *decrease* u_2.

But such problems do not arise when this inverse is representable as in (3.3.36). In the present case, observe that since $\lambda_W = \sqrt{2} = 1.414$ for this W matrix, it follows that if $|\rho| < 1/\lambda_W = 0.707$, then (3.3.36) must hold. But in case, the nonnegativity of ρ ensures that every term of the expansion, $\Sigma_{k=0}^{\infty} \rho^k W^k$, must be nonnegative, so that $(I_n - \rho W)^{-1}$ is *always nonnegative*. For example, if $\rho = .5$ then it can again be verified that

$$(3.3.39) \qquad (I_3 - \rho W)^{-1} = \begin{pmatrix} 1.5 & 1 & .5 \\ 1 & 2 & 1 \\ .5 & 1 & 1.5 \end{pmatrix}$$

So positive spatial dependencies here imply that spatial autoregressive residuals, u, are *always monotone nondecreasing* in the components of ε.

Finally, it should be emphasized that the negative signs in (3.3.37) are no accident. In fact it is shown in Section A3.5.3 of the Appendix that all elements of $(I_n - \rho W)^{-1}$ are nonnegative *if and only if* $|\rho| < 1/\lambda_W$. So while the steady-state representation in (3.3.36) is not strictly necessary for the existence of a reduced form solution for *u*, it characterizes those cases where a meaningful spatial interpretation of these residuals can be given.