7. Spatial Regression Parameter Estimation

Recall from the specification of both *SEM* in (6.1.3) and *SLM* in (6.2.2) above that the parameters, (β,σ^2,ρ), are essentially the same for both. As mentioned already, the key difference is how the spatial autoregressive hypothesis is applied (namely to the unobserved errors in *SEM* and to the observed dependent variable itself in *SLM*). So it is not surprising that the method of estimation is very similar for both of these models. But unlike the *iterative* estimation scheme employed for geo-kriging models in Section 7.3.1 of Part II (based on iteratively reweighted least-squares), the present method involves the *simultaneous* estimation of all model parameters. So our first objective is to develop this general method of *maximum-likelihood estimation*, and then to apply this method to both *SEM* and *SLM*.

7.1 The Method of Maximum-Likelihood Estimation

While maximum-likelihood estimation can in principle be applied to estimate the parameters of *any* probability model, it should be clear that the models of primary interest for our purposes are all based on the multi-normal model. So the following development is restricted to such models. Here the basic idea can be motivated by the following (extremely simplified) estimation problem for normal distributions. Suppose that a single sample, Y, is drawn from one of two possible populations having normal densities, ϕ_1 and ϕ_2, [as in expression (3.1.10) of Part II] with common unit variance, but with different means, $\mu_1 = 0$, and $\mu_2 = 2$. Here the problem is to estimate the true value of the mean based on the value, $Y = y$, of this one observation, as shown in Figure 7.1 below:

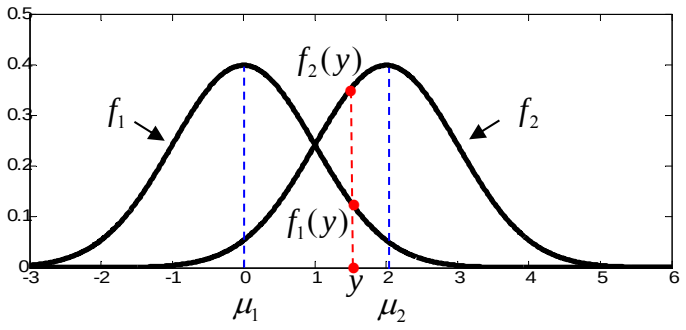

Figure 7.1. Simple Estimation Problem

To do so, observe that while the density values, $f_1(y)$ and $f_2(y)$, are not themselves probabilities, their *ratio* is approximately the *relative likelihood* of observing values from these two populations in any sufficiently small neighborhood, $[y-\varepsilon, y+\varepsilon]$, of y, as

shown in Figure 7.2 below. In particular, the area under each density, f_i, is seen to be well approximated by a rectangle with base length, 2ε, and height, $f_i(y)$, $i = 1,2$.

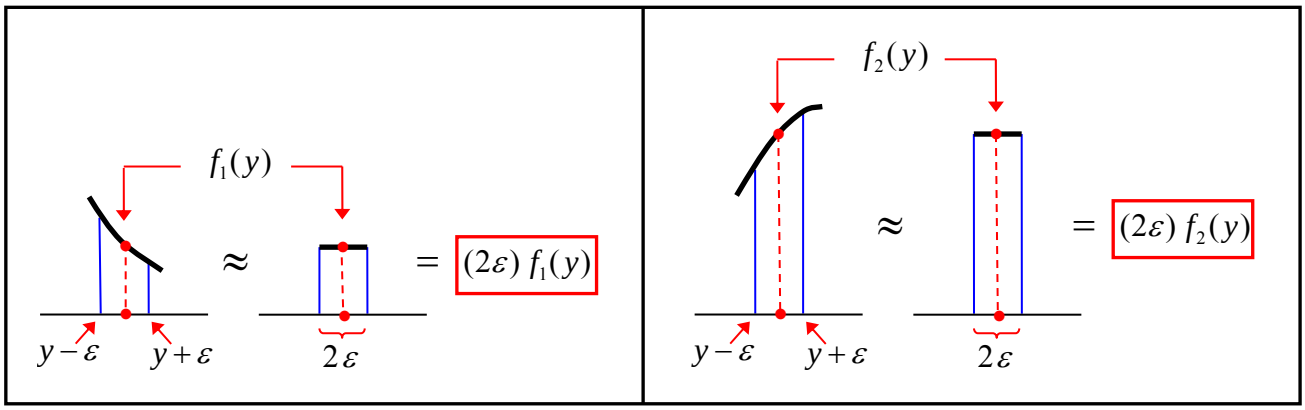

Figure 7.2. Relation of Density to Local Occupancy Probabilities

This figure shows that for any sufficiently small positive increment, ε,

$$(7.1.1) \qquad \frac{\Pr(y-\varepsilon \le Y \le y+\varepsilon \,|\, \mu_1)}{\Pr(y-\varepsilon \le Y \le y+\varepsilon \,|\, \mu_2)} \approx \frac{f_1(y)}{f_2(y)}$$

So if $f_2(y) > f_1(y)$, as in the present example, then it is reasonable to infer that y is more likely to have come from population 2 than population 1. More formally, we now say the *maximum-likelihood estimate*, $\hat{\mu}$, of the unknown mean μ, in this two-population case is given by:

$$(7.1.2) \qquad \hat{\mu} = \mu_i \;\Leftrightarrow\; f_i(y) > f_j(y)\,, \quad i,j \in \{1,2\},\, i \ne j$$

Next suppose that nothing is known about the mean of this population, so that μ could in principle be any real value. In this case, there is a continuum of possible normal populations, $\{f(\cdot \,|\, \mu),\, \mu \in \mathbb{R}\}$, to be considered. But it should still be clear that y is most likely to have come from that population for which the probability density, $f(y \,|\, \mu)$, is largest. Thus the *maximum-likelihood estimate*, $\hat{\mu}$, is now given by the condition that,

$$(7.1.3) \qquad f(y \,|\, \hat{\mu}) \;=\; \max_{\mu \in \mathbb{R}} f(y \,|\, \mu)$$

More generally, suppose we consider a given sample, $y_0 = (y_{01},..,y_{0n})$, of a random vector, $Y = (Y_1,..,Y_n)$, with multi-normal density, $f(y \,|\, \theta)$, where $\theta = (\theta_1,..,\theta_k)$ denotes the vector of relevant parameters defining this density. Here, θ, could in principle

contain all mean parameters, $\mu = (\mu_1,..,\mu_n)$, together with all covariance parameters, $\Sigma = (\sigma_{ij} : i, j = 1,..,n)$ defining f [as in expression (3.2.11) of Part II]. But more typically, θ, will contain a much smaller set of parameters that are assumed to completely specify both μ and Σ in any given model (as will be illustrated by the many examples to follow). Even in this general setting, the above notion of maximum-likelihood estimator continues to be perfectly meaningful. For example, suppose that $n = 2$, so that each candidate population is representable by a bivariate normal density similar to that Figure 3.2 of Part II. Then as a two-dimensional analogue to Figure 7.2 above, one can imagine the portion of density above a small rectangular neighborhood of $y_0 = (y_{01}, y_{02})$, as shown schematically on the left side of Figure 7.3 below.

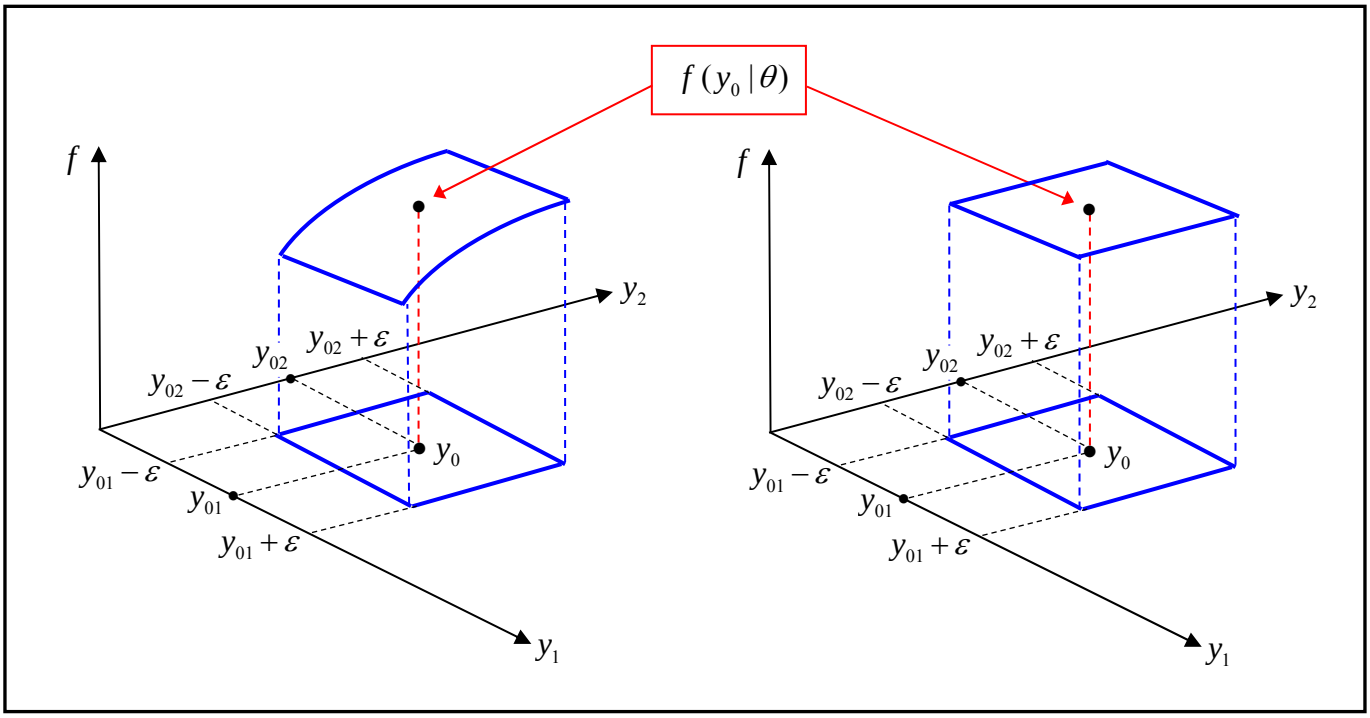

Figure 7.3. Local Occupancy Probabilities for Bivariate Densities

Here again, it is clear that for sufficiently small positive increments, ε, this density volume is well approximated by the box with base area, $(2\varepsilon)^2$, and height, $f(y_0, \theta)$, so that for any candidate parameter vectors, θ_1 and θ_2, we again have the approximation[1]

$$(7.1.4) \qquad \frac{\Pr(y_0 - \varepsilon 1_2 \le Y \le y + \varepsilon 1_2 \mid \theta_1)}{\Pr(y_0 - \varepsilon 1_2 \le Y \le y_0 + \varepsilon 1_2 \mid \theta_2)} \approx \frac{f(y_0 \mid \theta_1)}{f(y_0 \mid \theta_2)}$$

[1] Recall that 1_n is the unit vector in $\mathbb{R}^n$.

While such graphic representations are not possible in higher dimensions, $n > 2$, it should be clear that the same approximations hold for all n. So as a direct extension of (7.1.3), it follows that for any given sample observation, $y \in \mathbb{R}^n$, if the relevant set of possible values of a given parameter vector, $\theta = (\theta_1,..,\theta_k)$, is denoted by $\Theta \subseteq \mathbb{R}^k$,[2] then the *maximum-likelihood estimate*, $\hat{\theta}$, of parameter vector θ is again defined by the condition that:

(7.1.5) $$f(y \mid \hat{\theta}) = \max_{\theta \in \mathbb{R}^k} f(y \mid \theta)$$

Given the fact that sample y is the *known* quantity and θ is *unknown*, it is usually more convenient to define the corresponding *likelihood function,* $l(\cdot \mid y)$, by

(7.1.6) $$l(\theta \mid y) \equiv f(y \mid \theta) \quad , \quad \theta \in \mathbb{R}^k$$

and replace condition (7.1.5) by

(7.1.7) $$l(\hat{\theta} \mid y) = \max_{\theta \in \mathbb{R}^k} l(\theta \mid y)$$

Finally, because densities are *positive* (in the range of realizable samples, y), and because the *log-likelihood function*,[3]

(7.1.8) $$L(\theta \mid y) \equiv \log[l(\theta \mid y)]$$

is always *monotone increasing* in $l(\theta \mid y)$, it follows that *maximum-likelihood estimates*, θ, can be equivalently characterized by the *log-likelihood* condition:

(7.1.9) $$L(\hat{\theta} \mid y) = \max_{\theta \in \mathbb{R}^k} L(\theta \mid y)$$

The reason for this transformation is that multivariate density functions often involve products – as exemplified by the important case of independent random sampling, $f(y \mid \theta) = f(y_1,..,y_n \mid \theta) = \Pi_{i=1}^n f(y_i \mid \theta)$. Moreover, since logs convert products to sums, this representation is often simpler to analyze (as for example when differentiating likelihood functions).

7.2 Maximum Likelihood Estimation for General Linear Regression Models

To apply this estimation procedure, we start in Section 7.2.1 by considering the most familiar case of Ordinary Least Squares (OLS). By applying essentially the same arguments as in Section 7.1 of Part II, we then extend these results to Generalized Least

[2] For example, if $\theta = (\theta_1, \theta_2) = (\mu, \sigma^2)$ for $N(\mu, \sigma^2)$, then $\Theta = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_2 \geq 0\}$.

[3] In these notes "log" always means *natural log*, so the symbols *ln* and *log* may be used interchangeably.

Squares (GLS) in Section 7.2.2. These maximum-likelihood estimates for GLS will then serve as the general framework for obtaining comparable results for the spatial regression models, SEM and SLM in Sections 7.3 and 7.4 below.

7.2.1 Maximum Likelihood Estimation for OLS

Here we start with the *standard linear model*,

(7.2.1) $$Y = X\beta + \varepsilon\ ,\ \ \varepsilon \sim N(0, \sigma^2 I_n)$$

which in turn implies that Y must be multi-normally distributed as, $Y = N(X\beta, \sigma^2 I_n)$. So as the special case of expression (3.2.11) in Part II with $\mu = X\beta$ and $\Sigma = \sigma^2 I_n$, it follows that Y has multi-normal density, $f(y \mid \beta, \sigma^2)$, given by

(7.2.2) $$f(y \mid \beta, \sigma^2) \ = \ (2\pi)^{-n/2} \mid \sigma^2 I_n \mid^{-1/2} e^{-\frac{1}{2}(y - X\beta)'(\sigma^2 I_n)^{-1}(y - X\beta)}$$

[where the parameter vector, θ, for the general version in (7.1.5) above is here given by $\theta = (\beta, \sigma^2) = (\beta_0, \beta_1, .., \beta_k, \sigma^2)$]. By observing that $\mid \sigma^2 I_n \mid^{-1/2} = (\sigma^{2n})^{-1/2} \mid I_n \mid = (\sigma^2)^{-n/2}$ and $(\sigma^2 I_n)^{-1} = \sigma^{-2} I_n^{-1} = \sigma^{-2} I_n$, we see that this density can be simplified to:

(7.2.3) $$f(y \mid \beta, \sigma^2) \ = \ (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)}$$

so that the appropriate log-likelihood function for the OLS model is given by:

(7.2.4) $$L(\beta, \sigma^2 \mid y) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

Thus to obtain the maximum-likelihood estimates, $(\hat{\beta}, \hat{\sigma}^2)$, of the model parameters, we must maximize (7.2.4) with respect to β and σ^2. To do so, notice first that since β appears only in the last term (which is negative), it follows that for any choice of σ^2, the function L is always maximized with respect to β by minimizing the *squared deviation function*:

(7.2.5) $$SSD(\beta) = (y - X\beta)'(y - X\beta)$$

in a manner identical to expressions (7.1.10) and (7.1.11) in Part II. Thus expression (7.1.12) of Part II shows that this solution is again given by:

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

While this simple identity might appear to suggest that there is really no need for maximum likelihood estimation in the case of OLS, the real power of this method

becomes evident when we turn to the estimation of σ^2. Indeed, the method of least squares used for OLS is not directly extendable to σ^2, so that other methods must be employed. Even in the case of geostatistical regression, where a comparable estimate of σ^2 was developed in expression (7.3.19) of Part II, the actual estimation procedure involved a rather *ad hoc* application of nonlinear least-squares procedure for fitting spherical variograms to data. But in the present setting, we can now obtain a theoretically more meaningful estimate. In particular, by substituting $\hat{\beta}$ from (7.2.5) into (7.2.4), we can derive the exact *maximum-likelihood estimate*, $\hat{\sigma}^2$, of σ^2 by minimizing the reduced function,

$$(7.2.7) \qquad L_c(\sigma^2 \mid y) \equiv L(\hat{\beta}, \sigma^2 \mid y)$$

$$= -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2\sigma^2}(y - X\hat{\beta})'(y - X\hat{\beta})$$

where the subscript "*c*" reflects the common designation of this function as the *concentrated likelihood function* of parameter, σ^2 [also called a *profile* likelihood function]. But since the first order condition for a maximum yields:

$$(7.2.8) \qquad 0 = \frac{d}{d\sigma^2} L_c(\sigma^2 \mid y) = -\tfrac{n}{2}\left(\tfrac{1}{\sigma^2}\right) + \tfrac{1}{2}\left(\tfrac{1}{(\sigma^2)^2}\right)(y - X\hat{\beta})'(y - X\hat{\beta})$$

$$= -n + \tfrac{1}{\sigma^2}(y - X\hat{\beta})'(y - X\hat{\beta})$$

we see that the *maximum-likelihood estimate* for σ^2 is given by,[4]

$$(7.2.9) \qquad \boxed{\hat{\sigma}^2 = \tfrac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta})}$$

This can be given a more familiar form in terms of estimated residuals, $\hat{\varepsilon} = (\hat{\varepsilon}_1, .., \hat{\varepsilon}_n)'$ as

$$(7.2.10) \qquad \hat{\sigma}^2 = \tfrac{1}{n}\hat{\varepsilon}'\hat{\varepsilon} = \tfrac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

which is seen to be the "natural" estimator of $\sigma^2 = \text{var}(\varepsilon) = E(\varepsilon^2)$.

7.2.2 Maximum Likelihood Estimation for GLS

To extend these estimation results to GLS, we start with the *general linear model*,

$$(7.2.11) \qquad \boxed{Y = X\beta + \varepsilon, \;\; \varepsilon \sim N(0, \sigma^2 V)}$$

[4] One may also check that the second derivative of L_c evaluated at $\hat{\sigma}^2$ is negative and thus yields a maximum.

where the matrix, V, is assumed to be *known*.[5] So in this setting, OLS is seen to be the special case with $V = I_n$. The key feature of this model is that, like the OLS model in (7.2.3) above, the only unknown parameters are the beta coefficients, β, together with the positive variance parameter, σ^2 [so that again, $\theta = (\beta, \sigma^2)$]. As with OLS, this implies that Y is again *multi-normally distributed*, where in this case, $Y \sim N(X\beta, \sigma^2 V)$, with density:

$$(7.2.12) \qquad f(y \mid \beta, \sigma^2) = (2\pi)^{-n/2} \mid \sigma^2 V \mid^{-1/2} e^{-\frac{1}{2}(y - X\beta)'(\sigma^2 V)^{-1}(y - X\beta)}$$

By employing parallel matrix identities, $\mid \sigma^2 V \mid^{-1/2} = (\sigma^{2n})^{-1/2} \mid V \mid^{-1/2} = (\sigma^2)^{-n/2} \mid V \mid^{-1/2}$ and $(\sigma^2 V)^{-1} = \sigma^{-2} V^{-1}$, this can again be simplified to:

$$(7.2.13) \qquad f(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \mid V \mid^{-1/2} e^{-\frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta)}$$

which is seen to yield the associated *log-likelihood function*:

$$(7.2.14) \qquad \boxed{L(\beta, \sigma^2 \mid y) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2}\log \mid V \mid - \frac{1}{2\sigma^2}(y - X\beta)'V^{-1}(y - X\beta)}$$

So to obtain the maximum-likelihood estimate, $\hat{\beta}$, of β, it now follows (as an extension of the OLS case) that for any choice of σ^2, the function L will be maximized by choosing $\hat{\beta}$ to minimize the quadratic form, $(y - X\beta)'V^{-1}(y - X\beta)$, [which is identical in form to expression (7.1.27) of Part II, and may again be interpreted as a type of *weighted least-squares* problem]. But at this point, we may now observe [as in expression (7.1.15) of Part II] that if T denotes the *Cholesky matrix* for V,[6] then the matrix identity

$$(7.2.15) \qquad V = TT' \;\Rightarrow\; V^{-1} = (T')^{-1}T^{-1} = (T^{-1})'T^{-1},$$

allows us to reduce this quadratic form as follows:

$$(7.2.16) \qquad \begin{aligned} (y - X\beta)'V^{-1}(y - X\beta) &= (y - X\beta)'(T^{-1})'T^{-1}(y - X\beta) \\ &= (T^{-1}y - T^{-1}X\beta)'(T^{-1}y - T^{-1}X\beta) \\ &= (\tilde{y} - \tilde{X}\beta)'(\tilde{y} - \tilde{X}\beta) \end{aligned}$$

But this is precisely the squared deviation function in (7.2.5) for the new data set, $\tilde{y} = T^{-1}y$ and $\tilde{X} = T^{-1}X$. So it follows at once from (7.2.6) that the GLS maximum-likelihood estimate, $\hat{\beta}$, of β is given [as in expressions (7.1.21) through (7.1.24) in Part II] by

[5] Unlike the model specification in expression (7.1.8) of Part II, the matrix V need *not* be a correlation matrix (i.e., its diagonal elements need not be all ones). However, since $\sigma^2 V$ is required to be a *nonsingular covariance matrix*, V, must be *symmetric* and *positive definite* (as in Section A2.7.2 of the Appendix to Part II).

[6] Here existence of T is ensured by the *Cholesky Theorem* in Section A2.7.2 of the Appendix to Part II.

(7.2.17) $$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} = [(T^{-1}X)'(T^{-1}X)]^{-1}(T^{-1}X)'(T^{-1}y)$$
$$= [\hat{X}(T')^{-1}T^{-1}X]^{-1}X'(T')^{-1}T^{-1}y$$

so that by (7.2.15),

(7.2.18) $$\boxed{\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y}$$

Moreover, precisely the same maximization arguments for σ^2 in (7.2.8) and (7.2.9) above now show that the GLS *maximum-likelihood estimate* for σ^2 is given by

(7.2.19) $$\hat{\sigma}^2 = \tfrac{1}{n}(\tilde{y} - \tilde{X}\hat{\beta})'(\tilde{y} - \tilde{X}\hat{\beta}) = \tfrac{1}{n}(T^{-1}y - T^{-1}X\hat{\beta})'(T^{-1}y - T^{-1}X\hat{\beta})$$
$$= \tfrac{1}{n}(y - X\hat{\beta})'(T^{-1})'(T^{-1})(y - X\hat{\beta})$$

so that again by (7.2.15),

(7.2.20) $$\boxed{\hat{\sigma}^2 = \tfrac{1}{n}(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})}$$

Thus the maximum-likelihood estimation results for OLS are seen to be directly extendable to the class of GLS models (7.2.11).

7.3 Maximum Likelihood Estimation for SEM

To apply these general results to SE-models, we start by recalling from expressions (6.1.7) and (6.1.8) that SEM can be written as

(7.3.1) $$Y = X\beta + u, \quad u \sim N(0, \sigma^2 V_\rho)$$

where the *spatial covariance structure*, V_ρ, is given by

(7.3.2) $$V_\rho = (B_\rho' B_\rho)^{-1} = B_\rho^{-1}(B_\rho^{-1})'$$

with B_ρ given in terms of weight matrix, *W*, by

(7.3.3) $$B_\rho = I_n - \rho W$$

So SEM can be viewed as an instance of the GLS model in (7.2.11), where *V* now takes the specific form V_ρ in (7.3.2). However, it must be emphasized that unlike (7.2.11), the matrix V_ρ involves an *unknown parameter*, ρ. So to be precise, (7.3.1) should be viewed

as a GLS model *conditioned on* ρ. But nonetheless, we can still employ (7.2.14) to write down the appropriate *log-likelihood function* for SEM as

(7.3.4) $$L(\beta,\sigma^2,\rho\,|\,y) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\sigma^2) - \tfrac{1}{2}\log|V_\rho| - \tfrac{1}{2\sigma^2}(y - X\beta)'V_\rho^{-1}(y - X\beta)$$

In particular, we now know from (7.2.18) and (7.2.20) that for any given value of ρ, the maximum-likelihood estimates for β and σ^2, *conditional on* ρ, are given respectively by

(7.3.5) $$\hat{\beta}_\rho = (X'V_\rho^{-1}X)^{-1}X'V_\rho^{-1}y$$

and

(7.3.6) $$\hat{\sigma}^2_\rho = \tfrac{1}{n}(y - X\hat{\beta}_\rho)'V_\rho^{-1}(y - X\hat{\beta}_\rho)$$

where the subscript on these estimates reflects their dependency on the value of ρ. But since these conditional estimates are expressible as explicit (closed form) functions of ρ, we can substitute these results into (7.3.4) and obtain a *concentrated likelihood function* for ρ in a manner similar to that of σ^2 in the case of OLS [in expression (7.2.7) above]. In the present case, this concentrated likelihood takes the following form:

(7.3.7) $$L_c(\rho\,|\,y) \equiv L(\hat{\beta}_\rho,\hat{\sigma}^2_\rho,\rho)$$

$$= -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\hat{\sigma}^2_\rho) - \tfrac{1}{2}\log|V_\rho| - \tfrac{1}{2\hat{\sigma}^2_\rho}(y - X\hat{\beta}_\rho)'V_\rho^{-1}(y - X\hat{\beta}_\rho)$$

To further simplify this expression, we first note from (7.3.6) that the last term in (7.3.7) reduces to a constant, since

(7.3.8) $$-\tfrac{1}{2\hat{\sigma}^2_\rho}(y - X\hat{\beta}_\rho)'V_\rho^{-1}(y - X\hat{\beta}_\rho) = -\tfrac{1}{2\hat{\sigma}^2_\rho}[n\cdot\hat{\sigma}^2_\rho] = -\tfrac{n}{2}$$

Moreover, it follows from standard properties of matrix inverses and determinants [as in expressions (A31.18), (A3.1.20), (A3.2.70) and (A3.2.71) of the Appendix] that

(7.3.9) $$|V_\rho| = |(B'_\rho B_\rho)^{-1}| = |B_\rho^{-1}|\cdot|(B'_\rho)^{-1}| = |B_\rho|^{-1}|B'_\rho|^{-1} = |B_\rho|^{-2}$$

So by substituting these identities into (7.3.7) we obtain the simpler form of the concentrated likelihood function for ρ:

(7.3.10) $$L_c(\rho\,|\,y) = = -\tfrac{n}{2}[1 + \log(2\pi)] + \log|B_\rho| - \tfrac{n}{2}\log(\hat{\sigma}^2_\rho)$$

With these results, the desired maximum-likelihood estimation procedure for SEM is now evident. In particular, we first maximize the concentrated likelihood function, $L_c(\rho \mid y)$, to obtain the estimate, $\hat{\rho}$, and then use (7.3.5) and (7.3.6) to obtain the remaining estimates, $\hat{\beta}$ and $\hat{\sigma}^2$ as:

(7.3.11) $$\hat{\beta} = \hat{\beta}_{\hat{\rho}} = (X'V_{\hat{\rho}}^{-1}X)^{-1}X'V_{\hat{\rho}}^{-1}y$$

and

(7.3.12) $$\hat{\sigma}^2 = \hat{\sigma}^2_{\hat{\rho}} = \tfrac{1}{n}(y - X\hat{\beta})'V_{\hat{\rho}}^{-1}(y - X\hat{\beta})$$

Since $L_c(\rho \mid y)$ is a smooth function in one variable, the first step can be accomplished by standard numerical "line search" methods. So for reasonably small sample sizes, *n*, this estimation procedure is very efficient.

But for larger sample sizes (say, $n > 500$), an additional problem is created by the need to evaluate the *n*-square determinant, $|B_\rho|$, at each step of this procedure. However, such computations can often be made more efficient by means of the following observation. Recall from the discussion of eigenvalues and eigenvectors in Section 3.3.1 above that nonsingular matrices such as B_ρ have a "spectral" representation in terms of the diagonal matrix, $\Lambda_\rho = diag(\lambda_{\rho 1},..,\lambda_{\rho n})$, of their eigenvalues, together with the nonsingular matrix, $X_\rho = (x_{\rho 1},..,x_{\rho n})$, of their associated eigenvectors as:

(7.3.13) $$B_\rho = X_\rho \Lambda_\rho X_\rho^{-1}$$

So again by standard determinant identities [(A3.2.70) and (A3.2.72) in the Appendix], it follows that

(7.3.14) $$|B_\rho| = |X_\rho| \cdot |\Lambda_\rho| \cdot |X_\rho^{-1}| = |X_\rho| \cdot |\Lambda_\rho| \cdot |X_\rho|^{-1} = |\Lambda_\rho| = \prod_{i=1}^{n} \lambda_{\rho i}$$

Moreover, if the eigenvalues of the weight matrix, W, in (7.3.3) are denoted by λ_i with associated eigenvectors, x_i, $i = 1,..,n$, so that

(7.3.15) $$W x_i = \lambda_i x_i \quad , \quad i = 1,..,n$$

then it follows from (7.3.3) that

(7.3.16) $$B_\rho x_i = (I_n - \rho W)x_i = x_i - \rho W x_i = x_i - \rho\lambda_i x_i = (1 - \rho\lambda_i)x_i \quad , \quad i = 1,..,n$$

Thus we see that the eigenvalues of B_ρ are obtainable from those of W by the identity

(7.3.17) $\lambda_{\rho i} = 1 - \rho\lambda_i \quad , \quad i = 1,..,n$

(with corresponding eigenvector, $x_{\rho i} \equiv x_i$). In particular, this implies from (7.3.13) that

(7.3.18) $|B_\rho| = \prod_{i=1}^{n}(1 - \rho\lambda_i)$

and thus that the *log determinant* in (7.3.10) is given simply by

(7.3.19) $\log|B_\rho| = \sum_{i=1}^{n}\log(1 - \rho\lambda_i)$

So by calculating the eigenvalues $(\lambda_1,..,\lambda_n)$ of the weight matrix, *W*, we can rapidly compute the determinant, $|B_\rho|$, for *any* value of ρ. While the computation of these eigenvalues can itself be time consuming, the key point is that this calculation need only be done *once*. This procedure is so useful, that it is incorporated into almost all software packages for calculating such maximum-likelihood estimates (when *n* is sufficiently large).[7]

7.4 Maximum-Likelihood Estimation for SLM

In most respects, maximum-likelihood estimation for SL-models is virtually identical to that for SE-models. To begin with, recall from expression (6.2.6) that SLM can be written as

(7.4.1) $Y = X_\rho\beta + u \quad , \quad u \sim N(0, \sigma^2 V_\rho)$

where $X_\rho = B_\rho^{-1}X$ and where V_ρ and B_ρ are again given by (7.3.2) and (7.3.3). So the only formal difference here is that for each given value of ρ, we now obtain a GLS model in which both V and X depend on ρ. So the corresponding log likelihood function takes the form,

(7.4.2) $L(\beta, \sigma^2, \rho \mid y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|V_\rho| - \frac{1}{2\sigma^2}(y - X_\rho\beta)'V_\rho^{-1}(y - X_\rho\beta)$

which in turn implies that for the SLM case, the maximum-likelihood estimate for β conditional on ρ is given by:

(7.4.3) $\hat{\beta}_\rho = (X_\rho' V_\rho^{-1} X_\rho)^{-1} X_\rho' V_\rho^{-1} y$

[7] However, it should also be noted that for extremely large sample sizes (say $n > 1000$) the numerical accuracy of such eigenvalue calculations becomes less reliable. In such cases, (7.3.19) is often approximated by using only those terms with eigenvalues of largest absolute magnitudes.

For computational purposes, it is often more convenient to reduce this expression by observing that

$$(7.4.4) \qquad \begin{aligned} X'_\rho V_\rho^{-1} X_\rho &= (B_\rho^{-1}X)'(B'_\rho B_\rho)B_\rho^{-1}X \\ &= X'(B_\rho^{-1})'(B'_\rho B_\rho)B_\rho^{-1}X \\ &= X'\left((B'_\rho)^{-1}B'_\rho\right)\left(B_\rho B_\rho^{-1}\right)X = X'X \end{aligned}$$

and similarly that

$$(7.4.5) \qquad X'_\rho V_\rho^{-1} = (B_\rho^{-1}X)'(B'_\rho B_\rho) = X'\left((B'_\rho)^{-1}B'_\rho\right)B_\rho = X'B_\rho$$

So the maximum-likelihood estimate of β given ρ reduces to the simpler form

$$(7.4.6) \qquad \boxed{\hat{\beta}_\rho = (X'X)^{-1}X'B_\rho y}$$

Similarly, the maximum-likelihood estimate for σ^2 conditional on ρ is given by

$$(7.4.7) \qquad \hat{\sigma}^2_\rho = \tfrac{1}{n}(y - X_\rho\hat{\beta}_\rho)'V_\rho^{-1}(y - X_\rho\hat{\beta}_\rho)$$

But by using the same arguments in (7.4.4) and (7.4.5) we see that

$$(7.4.8) \qquad \begin{aligned} (y - X_\rho\hat{\beta}_\rho)'V_\rho^{-1}(y - X_\rho\hat{\beta}_\rho) &= (y - B_\rho^{-1}X\hat{\beta}_\rho)'(B'_\rho B_\rho)(y - B_\rho^{-1}X\hat{\beta}_\rho) \\ &= [B_\rho^{-1}(B_\rho y - X\hat{\beta}_\rho)]'(B'_\rho B_\rho)[B_\rho^{-1}(B_\rho y - X\hat{\beta}_\rho)] \\ &= (B_\rho y - X\hat{\beta}_\rho)'[(B'_\rho)^{-1}B'_\rho](B_\rho B_\rho^{-1})(B_\rho y - X\hat{\beta}_\rho)] \\ &= (B_\rho y - X\hat{\beta}_\rho)'(B_\rho y - X\hat{\beta}_\rho) \end{aligned}$$

and thus that the maximum-likelihood estimate for σ^2 conditional on ρ for SLM reduces to:

$$(7.4.9) \qquad \boxed{\hat{\sigma}^2_\rho = \tfrac{1}{n}(B_\rho y - X\hat{\beta}_\rho)'(B_\rho y - X\hat{\beta}_\rho)}$$

By substituting these expressions into (7.4.2), we again obtain a concentrated log likelihood function for ρ, namely

$$(7.4.10) \qquad L_c(\rho\,|\,y) = -\tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\log(\hat{\sigma}^2_\rho) - \tfrac{1}{2}\log|V_\rho| - \tfrac{1}{2\hat{\sigma}^2_\rho}(y - X_\rho\hat{\beta}_\rho)'V_\rho^{-1}(y - X_\rho\hat{\beta}_\rho)$$

As with SEM, this can be reduced by again observing from (7.4.7) that

(7.4.11) $$-\tfrac{1}{2\hat{\sigma}^2_\rho}(y - X_\rho\hat{\beta}_\rho)'V_\rho^{-1}(y - X_\rho\hat{\beta}_\rho) = -\tfrac{1}{2\hat{\sigma}^2_\rho}n\hat{\sigma}^2_\rho = -\tfrac{n}{2}$$

which together with (7.3.9) shows that the concentrated likelihood function for ρ has *exactly the same form* for SLM and for SEM, i.e.,

(7.4.12) $$L_c(\rho \mid y) = = -\tfrac{n}{2}[1+\log(2\pi)] + \log|B_\rho| - \tfrac{n}{2}\log(\hat{\sigma}^2_\rho)$$

So the only difference between (7.3.10) and (7.4.12) is the explicit form of $\hat{\sigma}^2_\rho$ in (7.3.6) and (7.4.9), respectively. In particular, this implies that all the discussion about numerical maximization of concentrated likelihoods to obtain $\hat{\rho}$ is identical for both models. In particular, the eigenvalue decomposition in (7.3.19) is precisely the same. So to complete the estimation procedure, it remains only to substitute this estimate, $\hat{\rho}$, into (7.4.6) and (7.4.9) to obtain the respective estimates,

(7.4.13) $$\hat{\beta} = (X'X)^{-1}X'B_{\hat{\rho}}y$$

and

(7.4.14) $$\hat{\sigma}^2 = \tfrac{1}{n}(B_{\hat{\rho}}y - X\hat{\beta})'(B_{\hat{\rho}}y - X\hat{\beta})$$

7.5 An Application to the Irish Blood Group Data

At this point, it is instructive to consider an application of these spatial regression models to an empirical example, namely the *Irish Blood Group* data in Section 1.2 above. To do so, we start with a standard OLS regression analysis in Section 7.5.1 below and test the residuals for spatial autocorrelation (as in Section 4.3 above). The spatial regression models, SEM and SLM, are then applied to this data in Section 7.5.2.

7.5.1 OLS Residual Analysis and Choice of Spatial Weights Matrices

Recall from Figures 1.7 and 1.8 above that the "footprint" of the 12th Century Anglo-Norm counties, known as the *Pale*, can still be seen in the spatial density pattern of Blood Group A in 1958. So an interesting question to explore is how much of this pattern can be statistically accounted for by this single explanatory variable.[8] To do so, we now consider a simple regression

(7.5.1) $$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \;\;,\;\; i = 1,..,n$$

[8] Note that the Irish Blood Group data set in [BG] contains one other potentially relevant explanatory variable, namely the number of place names (per unit of area) ending in "town" within each county. However, in the present example we focus only on the (marginal) effect of the Pale itself.

where relevant dependent variable, Y_i is the proportion of adults with *Blood Group A* in each county i, and the single explanatory variable, x_i, is taken to be the indicator (zero-one) variable for the *Pale* (corresponding to the red area in Figure 1.8 above), where

$$x_i = \begin{cases} 1 & , if\ i \in Pale \\ 0 & , if\ i \notin Pale \end{cases} \qquad (7.5.2)$$

To run this regression, we here use the ARCMAP version of OLS, and employ the ARCMAP data set in **Eire.mxd**. While JMP is generally more suitable for such analyses, performing OLS inside ARCMAP has the particular advantage of allowing the regression residuals to be mapped directly. This program can be found on the **ArcToolbox** path:

Spatial Statistics Tools > Modeling Spatial Relationships > Ordinary Least Squares

In the window that opens, type the entries shown on the left in Figure 7.4 below (where as usual, path names are machine specific):

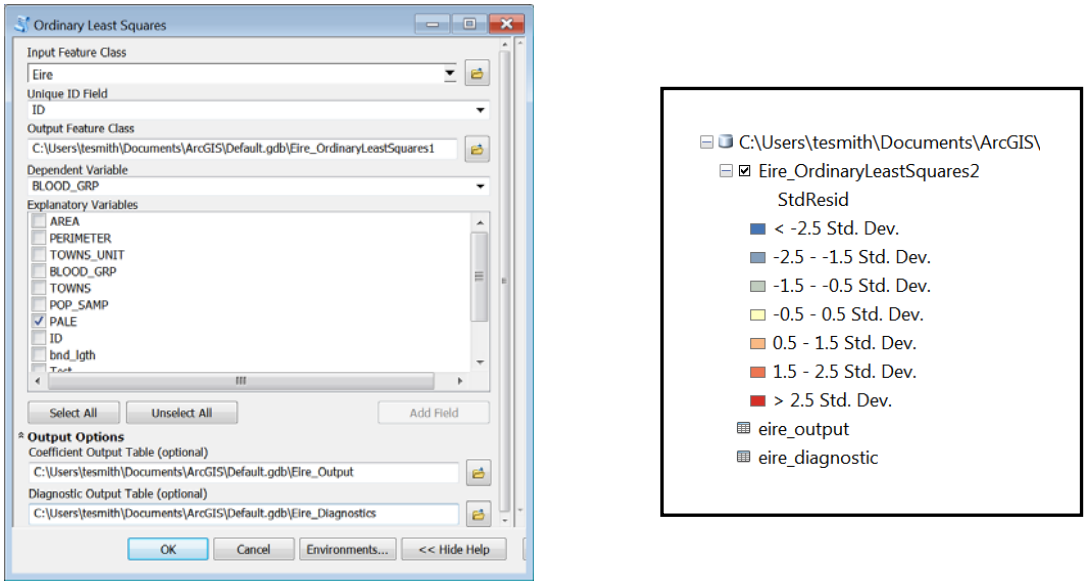

Figure 7.4. Running Ordinary Least Squares in ArcToolbox

Notice that both the *coefficient estimates* and *diagnostics* are "optional" tables, which should definitely be added. These will appear in the *Table of Contents*, as shown at the bottom right in Figure 7.4. The relevant portion of **eire_output** (for our purposes)[9] is shown in Table 7.1 below:

[9] Note in particular that the "robust" estimates and tests in this Table have not been shown. As with a number of other statistical diagnostics in ARCMAP, these robust-estimation results are difficult to interpret without further documentation.

Variable	Coef	StdError	t_Stat	Prob
Intercept	27.563571	0.528208	52.183184	0
PALE	4.252262	0.777501	5.46914	0.000012

Table 7.1. Coefficient Estimates and P-Values

So the "Pale effect" is seen to be positive and very significant, indicating that Blood Group A levels are significantly higher inside the Pale than elsewhere in Eire. But as we have seen many times before, this significance may well be inflated by the presence of spatial dependencies among Blood Group levels that are not accounted for by the Pale alone. So the remaining task is to test the *regression residuals* for spatial autocorrelation. These residuals are shown graphically on the right in Figure 7.6 below, where the pattern of Blood Group values in Figure 1.7 is reproduced on the left for ease of comparison.

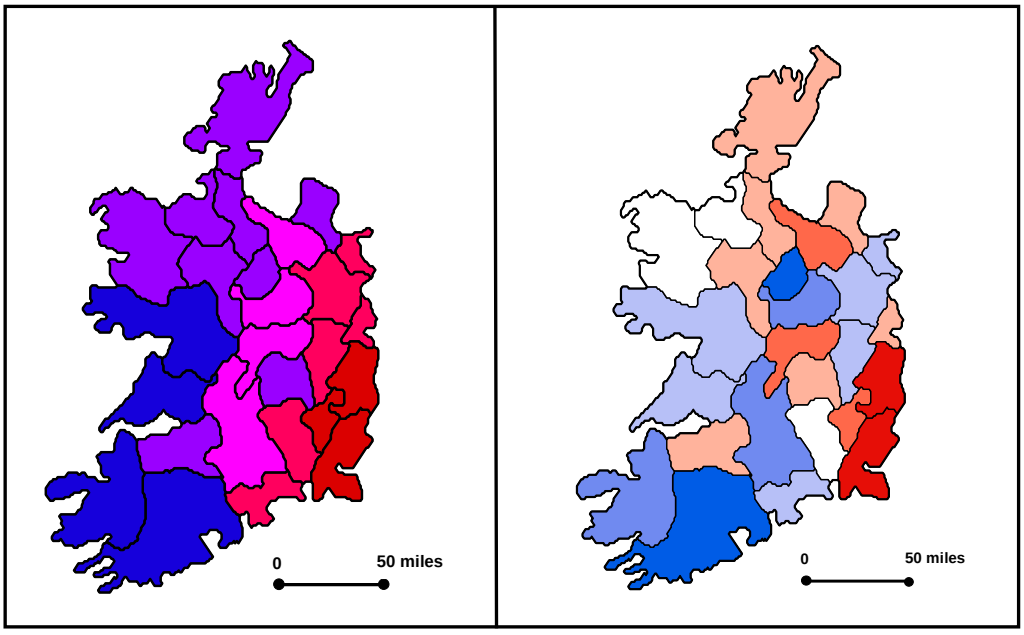

Figure 7.5. Blood Group values and OLS Residuals

Before analyzing these residuals, it is important to emphasize that the "default" residuals that appear in ARCMAP (as indicated on the right side of Figure 7.4) have been normalized to *Studentized Residuals* (**StdResiduals**). So to be comparable with the rest of our analysis, this plot must be redone in terms of the **Residuals** column in the *Attribute Table*, as is done in Figure 7.5.[10]

[10] Note that studentized residuals (again not documented in ARCMAP) are useful for many testing purposes when the original assumption of *independent* residuals holds. But in the presence of possible spatial dependencies, it is generally preferable to analyze the raw residuals themselves.

Note from the plot of these residuals in Figure 7.5 that (as with many linear regressions) the highest Blood Group values in the Pale are underestimated (red residuals) and the lowest values outside the Pale are overestimated (blue residuals). This by itself tends to preserve a certain amount of the positive correlation seen in the original Blood Group data.

But to determine the statistical significance of such residual correlations, we must of course employ an appropriate spatial weights matrix, *W*. Because the present Eire example provides a dramatic illustration of how important this choice of *W* can be, we now consider this choice in some detail. To do so, a number of candidate weight matrices from Sections 2.1.2 and 2.1.3 were applied to this residual data, with test results summarized in terms of p-values in Table 7.2 below. Here the first column **Asymp** displays the results of the standard *asymptotic Moran test* in Section 4.2.1 above. The remaining three colums, **Moran**, **Rho**, and **Corr**, show comparable results for the *sac_perm test* in Section 4.3.1 above (using 999 simulations).

	Asymp	Moran	Rho	Corr
W_nn	0.540	0.595	0.504	0.589
W_nns	0.235	0.282	0.280	0.285
W_nn5	0.228	0.113	0.117	0.115
W_queen	0.249	0.091	0.118	0.106
W_share	0.016	0.035	0.058	0.039
W	0.010	0.019	0.026	0.020

Table 7.2. P-values for the Eire OLS Residuals

The first spatial weights matrix considered is the simple (centroid) *nearest-neighbor* matrix, W_{nn}, which (as already mentioned above Figure 1.18) is very restrictive for areal data in terms of potentially relevant neighbors ignored. Here it is clear that no spatial autocorrelation is detected by any method using this matrix. A slightly more appropriate version is the *symmetric nearest-neighbor* matrix, W_{nns}, [expression (2.1.10) above with $k = 1$] shown in the next row. Here the results are all still very insignificant, but are nonetheless dramatically more significant than for the asymetric case. The reason for this in the case of Eire can be seen in Figure 7.6 below, where county centroids are shown as blue dots, and where the red line emanating from each centroid is directed toward its nearest neighbor. This figure (which extends the Laoghis County illustration in Figure 1.18 above) confirms that such neighbor relations are relatively sparse throughout Eire. In particular, there are very few *mutual* nearest neighbors, i.e., red lines with *both* ends connected to centroids. So when moving from nearest neighbors, W_{nn}, to symmetric nearest neighbors, W_{nns}, it is now clear that many more relations are added to the matrix, thus allowing many more possibilities for spatial correlation to be considered.

The third and fourth rows show respective results for the *queen contiguity* matrix, W_{queen}, [expression (2.1.15)] and for one of its *k*-nearest-neighbor approximations, namely, the *five nearest neighbor* version, W_{nn5} [as in expression (2.1.9) with $k = 5$, and as also used in Figure 1.18 for Laoghis County]. These two cases are of special interest, since they are by far the most commonly used weights matrices for analyzing areal data. But in both cases, spatial autocorrelation is at best seen to be weakly significant – and is totally insignificant for the standard asymptotic Moran test.[11]

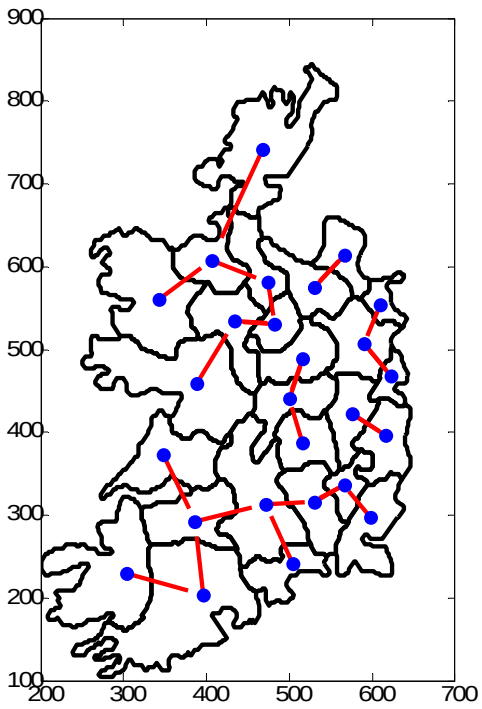

Figure 7.6. Nearest-Neighbor Relations in Eire

In view of this lack of significance, the results in the final two rows are quite striking. These show respective results for the *boundary shares* matrix, W_{share} [expression (2.1.17)], and for the *combined distance-shares* matrix, *W*, of Cliff and Ord (1969) [expression (2.1.18)]. Because we shall employ this latter matrix, *W*, in our subsequent analyses, it is here convenient to reproduce its typical elements, w_{ij}, as follows,

$$(7.5.3) \qquad w_{ij} = \frac{l_{ij} d_{ij}^{-1}}{\sum_{k \neq i} l_{ik} d_{ik}^{-1}}$$

[11] In fairness, it should be pointed out (as is done for example in ARCMAP) that such asymptotic tests typically require more samples (areal units) for statistical reliability. A common rule of thumb (that we have seen already for the Central Limit Theorem) is that *n* be at least 30.

where l_{ij} is the fraction of the boundary of county i shared with county j, and d_{ij} is the distance between their respective centroids.[12] While it is difficult to explain exactly why these two matrices capture so much more significance, one can gain insight by simply noting the unusual *complexity* of county boundaries in Eire. These complexities have most likely resulted from a long history of interactions between neighboring counties, so that shared boundary lengths may well reflect the degree of such interactions. Moreover, in so far as centroid distances tend to reflect relative travel distance between counties, it is reasonable to suppose that such distances reflect other dimensions of interaction. In any case, this example provides a clear case where it is prudent to consider a variety of tests in terms of alternative spatial weights matrices before drawing any firm conclusions about the presence of spatial autocorrelation. One rule of thumb is to try several (say three) different matrices which exhibit sufficient qualitative differences to capture a range of interaction possibilities. As stressed at the beginning of Part III, one of the most perplexing features of areal data analysis is the absence of any clear notion of "spatial separation" between areal units.

7.5.2 Spatial Regression Analyses

As stated above, we here employ the *combined distance-shares* matrix, *W*, in (7.5.3) which captures the most significant amount of spatial autocorrelation in Table 7.2 [and which constitutes the original matrix used by Cliff and Ord (1969) in their classic study of this Eire data]. To construct such a matrix, we first note that the procedure for constructing boundary-share weights is developed in Sections 3.2.2 and 3.2.3 of Part IV (as mentioned in Section 2.2.2 above), and is also discussed in more detail in Assignment 7. For the case of Eire, such boundary shares are given by matrix, **W_share**, in the MATLAB workspace, **Eire.mat**. Using the MATLAB script, **eire_wts.m**, these shares (l_{ij}) can be combined with centroid distances (d_{ij}) to yield the desired combined distance-shares weight matrix, **W**, in the workspace.[13]

Given this weight matrix, we now employ the spatial regression models, SEM and SAR, to capture the relation between Blood Group levels and the Pale in a manner that accounts for the spatial autocorrelation detected in Table 7.2. The estimation procedures for SEM and SLM are implemented in the MATLAB programs, **sem.m** and **slm.m**, respectively. The inputs required for each program consist of a *data vector*, **y**, for the dependent variable, a *data matrix*, **X**, for the explanatory variables, and an appropriate *spatial weights matrix*, **W**, relating the relevant set of areal units. In the present case, **y** is the vector of Blood Group proportions for each county, **X** is the vector, **x**, identifying those counties in the Pale, and **W** is the combined distance-shares matrix above.

[12] Further discussion of this weight matrix can be found in Upton and Fingleton (1985, pp.287-288) [see Reference 18 in the class *Reference Materials*].

[13] Here it is of interest to note that these weights differ slightly from those of Cliff and Ord (1969), which can be found in Table 5.1 of Upton and Fingleton (1985), and which are also reproduced as matrix, **W2**, in the workspace, **Eire.mat**. This illustrates the fact that such constructions will differ to some degree depending on the particular map of Eire that is used. (Indeed, digital maps did not even exist in 1969 when the original work was done.)

Before running these models, it should be noted that there are two additional inputs, **vnames** and **val** (also described in the program documentation). We have already seen **vnames** used as the list of variable names in previous applications (as for example in Cobalt Example of Section 7.3.4 in Part II). For the present case of a single variable, one need only write the variable name in single quotes, which here is **'Pale'**. The final input, **val**, represents the optional input of eigenvalues for **W** used to calculate the log determinant in (7.3.19) above. In the case of Eire with $n = 26$, this is hardly necessary. But for very large weight matrices, **W**, it is worth noting that the corresponding vector of eigenvalues is easily obtained in MATLAB with the command:

>> val = eig(W);

With these preliminary observations, we can now run both SEM and SLM, using the respective commands:

>> sem(y,X,W,'Pale');

>> slm(y,X,W,'Pale');

It should also be noted that there are a number of data outputs given by these two models. But for our present purposes, it is enough to examine their screen outputs, as shown in Figure 7.7 below. Here it is clear that there is a strong parallel between the output formats of each model. In particular, they are quite comparable in terms of both their output results and diagnostics (as discussed in more detail below). Note also that these two formats look very much the same as for OLS regression in the sense that significance levels (p-values) are reported for each parameter estimate, together with various measures of "goodness of fit". But as we shall see below, the actual methods of obtaining these results (and in some cases, even their meaning) differs substantially from OLS. Nonetheless, the basic interpretations of parameter estimates and their significance levels will remain the same as in OLS. So before getting into the details of calculation methods, it is appropriate to begin by examining these results in a qualitative way.

With respect to SEM, notice first that while the *Pale* effect continues to be positive (as in Table 7.1 for OLS), this effect is now both *smaller in magnitide* (1.55 versus 4.25) and dramatically *less significant* (with a p-value of .0788 versus .000012). Notice also that the level of spatial autocorrelation, $\hat{\rho} = 0.7885$, is significantly positive. As we have seen before, this suggests that such differences are largely due to the presence of spatial autocorrelation. While the exact nature of these effects is difficult to identify in the present spatial regression setting, we can nonetheless make certain useful observations. First, if the relevant data matrix for this Eire example is denoted by $X = [1_n, x]$, then it follows from expression (7.1.12) in Part II together with and (7.3.11) above that the OLS and SEM estimates of $\beta = (\beta_0, \beta_1)'$ are given respectively by

(7.5.4) $$\hat{\beta}_{OLS} = (X'X)^{-1}X'y$$

SEM OUTPUT

FINAL REGRESSION RESULTS:

VAR	COEFF	Z-VAL	PROB
const	28.82487	20.66107	0.000000
Pale	1.553209	1.757660	0.078805

Variance = 2.1251

AUTOCORRELATION RESULTS:

	VAL	Z-VAL	PROB
rho	0.788456	7.466704	0.000000

GOODNESS-OF-FIT RESULTS:

Extended R-Square = 0.3313
Extended R-Square Adj = 0.3034
Squared_Correlation = 0.5548
Log Likelihood Value = -49.8773
AIC = 107.7546
AIC_corrected = 109.6593
BIC = 112.7869

TESTS OF SEM MODEL:

TEST	VAL	PROB
LR	7.374837	0.006614
Com-LR	18.427035	0.000018

MORAN z-score and p-val = (0.2741,0.3920)

SLM OUTPUT

FINAL REGRESSION RESULTS:

VAR	COEFF	Z-VAL	PROB
const	7.130157	2.218746	0.026504
Pale	2.014177	3.471544	0.000517

Variance = 1.6146

AUTOCORRELATION RESULTS:

	VAL	Z-VAL	PROB
rho	0.726419	6.466525	0.000000

GOODNESS-OF-FIT RESULTS:

Extended R-Square = 0.7335
Extended R-Square Adj = 0.7224
Squared_Correlation = 0.7512
Log Likelihood Value = -45.6632
AIC = 99.3263
AIC_corrected = 101.2311
BIC = 104.3587

TEST OF SLM MODEL:

TEST	VAL	PROB
LR	15.803078	0.000070

MORAN z-score and p-val = (-0.7550,0.7749)

SAC_PERM TEST (N = 999)

INDEX	VALUE	SIGNIF
Moran	-0.0252	0.4544
corr	-0.0445	0.4534
rho	-0.0784	0.4541

SAC_PERM TEST (N = 999)

INDEX	VALUE	SIGNIF
Moran	-0.1734	0.8135
corr	-0.3110	0.8097
rho	-0.5579	0.8086

Figure 7.7. Regression Results and Autocorrelation Tests for SEM and SLM

and,

(7.5.5) $$\hat{\beta}_{SEM} = (X'V_{\hat{\rho}}^{-1}X)^{-1}X'V_{\hat{\rho}}^{-1}y$$

In contrast to OLS, the beta estimates for SEM are thus seen to depend on the estimated level of *spatial autocorrelation*, $\hat{\rho}$, together with the choice of *spatial weights matrix*, *W*, implicit in $V_{\hat{\rho}}$. So while in theory such estimates are still unbiased [recall expression (7.1.26) in Part II], their sensitivity to $\hat{\rho}$ tends to inflate the variance of these β estimates.

This can be seen in part by considering the standard errors of the estimated *Pale* parameter, $\hat{\beta}_1$, for both OLS and SEM. To do so, recall first from Table 7.1 that the standard error for $\hat{\beta}_1$ under OLS was given by,

(7.5.6) $$s_{OLS}(\hat{\beta}_1) = 0.7775$$

To derive the comparable standard error under SEM, we begin by noting that appropriate "Z-VAL" for $\hat{\beta}_1$ in Figure 7.7 is given [in a manner analogous to expression (7.3.26) in Part II] by

(7.5.7) $$z_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}},$$

so that the estimated standard error for $\hat{\beta}_1$ under SEM is given from Figure 7.7 by,

(7.5.8) $$s_{SEM}(\hat{\beta}_1) = \frac{\hat{\beta}_1}{z_{\hat{\beta}_1}} = \frac{1.553209}{1.757660} = 0.88368$$

This shows that standard errors of beta estimates do indeed tend to be larger in the presence of spatial autocorrelation.

Before turning to the SL-model, it is important to note that while the estimated spatial autocorrelation level, $\hat{\rho}$, for this SE-model is significantly positive, it is not evident that $\hat{\rho}$ has successfully eliminated *all* spatial autocorrelation effects found for weight matrix, *W*, in Table 7.2. To address this issue, we may again appeal to the results developed for all GLS models in expressions (7.1.18) and (7.1.19) in Part II, which show that if the spatial covariace structure, V_ρ, [in (7.3.2) and (7.3.3)] has been correctly estimated, then the Cholesky reduction of this model to *OLS form* should yield residuals that exhibit no significant spatial autocorrelation (with respect to *W*). In the present case, however, there is no need for Cholesky decompositions, since V_ρ in (7.3.2) is already factorized in terms

of B_ρ^{-1}. In fact the reduction of SEM to an OLS form can be made even more transparent by simply recalling from expression (6.1.9) that

(7.5.9) $$Y = X\beta + B_\rho^{-1}\varepsilon \ , \ \varepsilon \sim N(0,\sigma^2 I_n)$$

$$\Rightarrow \ B_\rho Y = B_\rho X\beta + \varepsilon \ , \ \varepsilon \sim N(0,\sigma^2 I_n)$$

$$\Rightarrow \boxed{Y_\rho = X_\rho\beta + \varepsilon \ , \ \varepsilon \sim N(0,\sigma^2 I_n)}$$

where $Y_\rho = B_\rho Y$ and $X_\rho = B_\rho X$. So to test the success of this SE-model it suffices to analyze the residuals:

(7.5.10) $$\hat{\varepsilon} = Y_{\hat{\rho}} - X_{\hat{\rho}}\hat{\beta}_{SEM}$$

of the estimated OLS model in (7.5.9), by again using **sac_perm.m**. Since this procedure is detailed in part (c) of Assignment 7, it suffices here to observe that the full command for **sem.m** in Section 7.5.2 above is of the form:

>> [OUT,cov,DAT] = sem(y,X,W,'Pale');

where the matrix **OUT** contains a number of useful transformations of the regression outputs. In particular, the residuals in (7.5.10) are contained in the third column, so that the command,

>> res_SEM = OUT(:,3);

produces a copy, **res_SEM**, of these residuals that can be tested using **sac_perm** as follows:

>> sac_perm(res_SEM,W,999);

The results of this test are shown in the lower left panel of Figure 7.7, and confirm that this application of SEM has indeed been successful in removing the spatial autocorrelation found under weight matrix, *W*.

Turning next to the SL-model, the most important difference to notice here is that while the *Pale* effect on Blood Group A is again positive – it is now *vastly* more significant than for the SE-model, with **p-value = 0.0005**. Moreover, by substituting the maximum-likelihood estimates $(\hat{\beta},\hat{\sigma}^2,\hat{\rho})$ for each model into their respective log-likehood functions in (7.3.4) and (7.4.2), we obtain maximum log-likehood values for SEM and SLM that constitute one possible measure of their *goodness of fit* to this Eire data (see Section 9 below for a more detailed discussion of goodness-of-fit measures). As seen in the GOODNESS-OF-FIT section for each model in Figure 7.7, these values are given respectively by,

(7.5.11) $$L_{SEM}(\hat{\beta},\hat{\sigma}^2,\hat{\rho}) = -49.8773$$

and

(7.5.12) $$L_{SLM}(\hat{\beta},\hat{\sigma}^2,\hat{\rho}) = -45.6632$$

So in terms of this likelihood comparison, it is clear that SLM also yields a *much better fit* to the Eire data than SEM (i.e., a much higher log-likelihood value).

This raises the natural question as to *why* SLM is so much more successful in capturing this spatial pattern of Blood Group A levels in Eire. Interestingly enough, the answer appears to lie in the *ripple effect* underlying the spatial autoregressive multiplier matrix, $B_\rho^{-1} = (I_n - \rho W)^{-1}$, for these models, as detailed in Section 3.3 above. The key point here is that while this ripple effect applies only to unobserved residuals in the SE-model, it also applies to the *explanatory variables* in the SL-model, as is evident in expression (6.2.4) above. More specifically, since our present weight matrix, *W*, in expression (7.5.3) is *row normalized*, it follows from expression (2.1.19) above that

(7.5.13) $$W1_n = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_j w_{1j} \\ \vdots \\ \sum_j w_{nj} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1_n$$

which in turn shows that

(7.5.14) $$B_\rho 1_n = (I_n - \rho W)1_n = 1_n - \rho W 1_n = (1-\rho)1_n$$

$$\Rightarrow 1_n = B_\rho^{-1}(B_\rho 1_n) = B_\rho^{-1}[(1-\rho)1_n] = (1-\rho)B_\rho^{-1}1_n$$

$$\Rightarrow B_\rho^{-1}1_n = \left(\tfrac{1}{1-\rho}\right)1_n$$

So in the present case, expression (6.2.4) for SL-models now takes the form:

(7.5.15) $$Y = B_\rho^{-1}[1_n, x]\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + B_\rho^{-1}\varepsilon = B_\rho^{-1}1_n\beta_0 + B_\rho^{-1}x\beta_1 + B_\rho^{-1}\varepsilon$$

$$= B_\rho^{-1}1_n\beta_0 + B_\rho^{-1}x\beta_1 + B_\rho^{-1}\varepsilon = 1_n\left(\tfrac{\beta_0}{1-\rho}\right) + (B_\rho^{-1}x)\beta_1 + B_\rho^{-1}\varepsilon$$

$$\Rightarrow \boxed{Y = \alpha_0 1_n + x_\rho \beta_1 + B_\rho^{-1}\varepsilon}$$

where $\alpha_0 = \beta_0/(1-\rho)$ and $x_\rho = B_\rho^{-1}x$. But since α_0 is essentially independent of ρ for estimation purposes (i.e., α_0 can assume any value given appropriate choices of β_0), it follows that the only difference between SL-model (7.5.15) and the SE-model in (7.5.9) is that *Pale* data vector, x, has now been tranformed to x_ρ. Moreover, recalling from expression (3.3.8) that x_ρ can be written as

$$(7.5.16) \qquad x_\rho = (I_n - \rho W)^{-1}x = x + \rho Wx + \rho^2 W^2 x + \cdots$$

it is natural to designate this transformed vector as the *rippled Pale*.

With these preliminary observartions, it should now be clear that the relative success of the SL-model versus the SE-model in this Eire case can be attributed entirely to this *rippled Pale effect*. The dramatic nature of this effect in the Eire case is illustrated in Figure 7.8 below, where values of the rippled Pale are plotted on the far right (and where the maximum and minumum values of the rippled Pale have been rescaled to be the same as those of Blood Group A). Further reflection suggests that this remarkable fit may not be simply a coincidence. Indeed, the gradual intermingling of blood-group types between Anglo-Normans and the indigenous Eire population might well be viewed as a "rippling" of intermarriage effects over many generations.

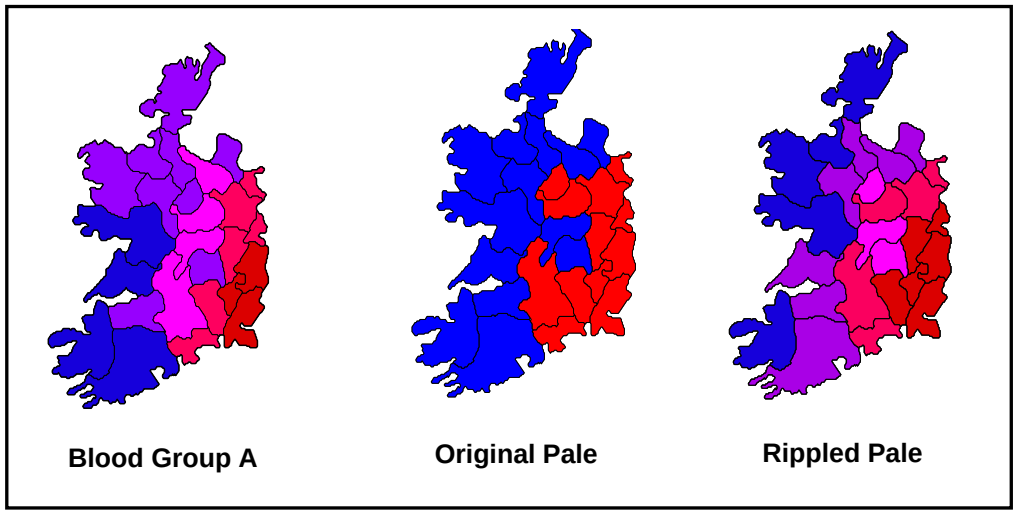

Figure 7.8. Comparison of Pale Effects and Rippled Pale Effects

With this qualitative overview of SEM and SLM applications to Eire, we turn now to a more detailed development of the many diagnostics displayed in Figure 7.7. To do so, we start in Section 8 below with a development of the fundamental *significance tests* for model parameters.