4. Testing for Spatial Autocorrelation

To apply the spatial autoregressive model above, we start by restating the *linear model* (for *n* areal units) in expression (3.2) above, where the *residuals*, *u*, are now specified more explicitly as:

(4.1) $\quad Y = X\beta + u \quad , \quad u \sim N(0,V)$

with both the parameter vector, β, and covariance matrix, V, unknown. The simplest procedure for specifying the residual covariance is to start by assuming that

(4.2) $\quad V = \sigma^2 I_n \quad ,$

so that β can be estimated by OLS. Given this estimate, one can then test to see whether there is sufficient *spatial autocorrelation* in the residuals to warrant more elaborate specifications of V. If for a given sample, y (i.e., observed realization of Y) we denote the *OLS estimate* of β by

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

and corresponding *OLS residuals* by

(4.4) $\quad \hat{u} = y - X\hat{\beta}$

then our objective is to develop statistical tests for the presence of *spatial autocorrelation* using these residuals. To do so, we assume that the underlying spatial structure of these *n* areal units is representable by a given *spatial weights matrix*, $W = (w_{ij} : i, j = 1,..,n)$. In terms of *W*, it is then hypothesized that all relevant spatial autocorrelation among the residuals, u, in (4.1) can be captured by the *spatial autoregressive model*,

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

The key feature of this hypothesis is that testing for spatial autocorrelation then reduces to testing the *null hypothesis*:

(4.6) $\quad H_0 : \rho = 0$

For if H_0 cannot be rejected, then (4.5) reduces to

(4.7) $\quad u = \varepsilon \sim N(0,\sigma^2 I_n) ,$

so that the OLS specification of V in (4.4) above is appropriate. If not, then some more elaborate specification of V needs to be considered.

4.1. Three Test Statistics

In this context, our main objective is to construct appropriate test statistics based on $\hat{u}$ and W for testing H_0. In the following subsections, we shall consider three alternative test statistics that are in common use.

4.1.1 Rho Statistic

Given model (4.5), one natural approach is simply to treat the OLS residuals, $\hat{u}$, as a sample of u, and use model (4.5) to obtain a corresponding OLS estimate of ρ. To do so, recall the "one variable regression" illustration given in class, where we started with a linear model:

(4.1.1) $Y_i = x_i b + \varepsilon_i$, $i = 1,..,n$

In vector form, this is seen to yield the special case of (4.4.1) where $X = x$ and $\beta = b$ is a scalar, i.e.,

(4.1.2) $Y = xb + \varepsilon$

Hence, as a special case of (4.3), the OLS estimate of the scalar, b, in (4.1.2) is given by

(4.1.3) $\hat{b} = (x'x)^{-1}x'y = \dfrac{x'y}{x'x} = \dfrac{x'y}{\|x\|^2}$

But (4.5) can be viewed as a model of the form (4.1.2), where $b = \rho$, $Y = u$ and $x = Wu$. Hence, for our present "data", $y = \hat{u}$, the corresponding OLS estimate of ρ is given by

(4.1.4) $\hat{\rho}_W = \dfrac{(W\hat{u})'\hat{u}}{(W\hat{u})'(W\hat{u})} = \dfrac{\hat{u}'W\hat{u}}{\hat{u}'W'W\hat{u}} = \dfrac{\hat{u}'W\hat{u}}{\|W\hat{u}\|^2}$

This yields our first test statistic for H_0, which we designate as the *rho statistic*. Note also that we use the subscript "W" to emphasize that this statistic (and those below) depends explicitly on the choice of W.

Having constructed this statistic, it is of interest to observe that the basic spatial autocorrelation test we have been using so far, namely regressing residuals on nearest-neighbor residuals, is essentially a special case of this rho statistic. To see this, observe that the ith row of (4.5) is of the form:

(4.1.5) $u_i = \rho(Wu)_i + \varepsilon_i = \rho\sum_{j=1}^{n} w_{ij}u_j + \varepsilon_i$

But if W is chosen to be the *first nearest-neighbor matrix* (i.e., $w_{ij} = 1$ if j is the nearest neighbor of i and $w_{ij} = 0$ otherwise) and if we let $nn(i)$ denote the *first nearest-neighbor* of each point i then by construction,

(4.1.6) $\sum_{j=1}^{n} w_{ij} u_j = u_{nn(i)}$

So for this case, (4.1.5) is of the form

(4.1.7) $u_i = \rho u_{nn(i)} + \varepsilon_i$

But this is almost exactly the regression we have been using, where the important slope coefficient is now precisely ρ. So the test for significance of this slope is based on the estimator, $\hat{\rho}_W$. Notice however that unlike our regression, there is *no intercept* in (4.1.7). This makes sense theoretically since[1]

(4.1.8) $E(u_i) = E(u_{nn(i)}) = E(\varepsilon_i) = 0$,

which in turn implies that the intercept term must also be *zero* in this model. So in fact, (4.1.7) is the model we *should* have been using. But since regression residuals must sum to zero by the construction of OLS,[2] the intercept is usually not statistically significant. This is well illustrated by the regression of Myocardial Infarctions on the Jarman Index in Section 1.3 above. Residual regression with and without the intercept are compared in Figures 4.1 and 4.2 below. Notice in Figure 4.1 that the intercept is close to zero and completely insignificant. More importantly, notice that the *t*-values for the slope intercepts in both figures are virtually identical.

Parameter Estimates (Figure 4.1)

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.0058051	0.015586	0.37	0.7100
nn-Resids	0.5111284	0.063697	8.02	<.0001*

Parameter Estimates (Figure 4.2)

Term	Estimate	Std Error	t Ratio	Prob>|t|
nn-Resids	0.5100272	0.063483	8.03	<.0001*
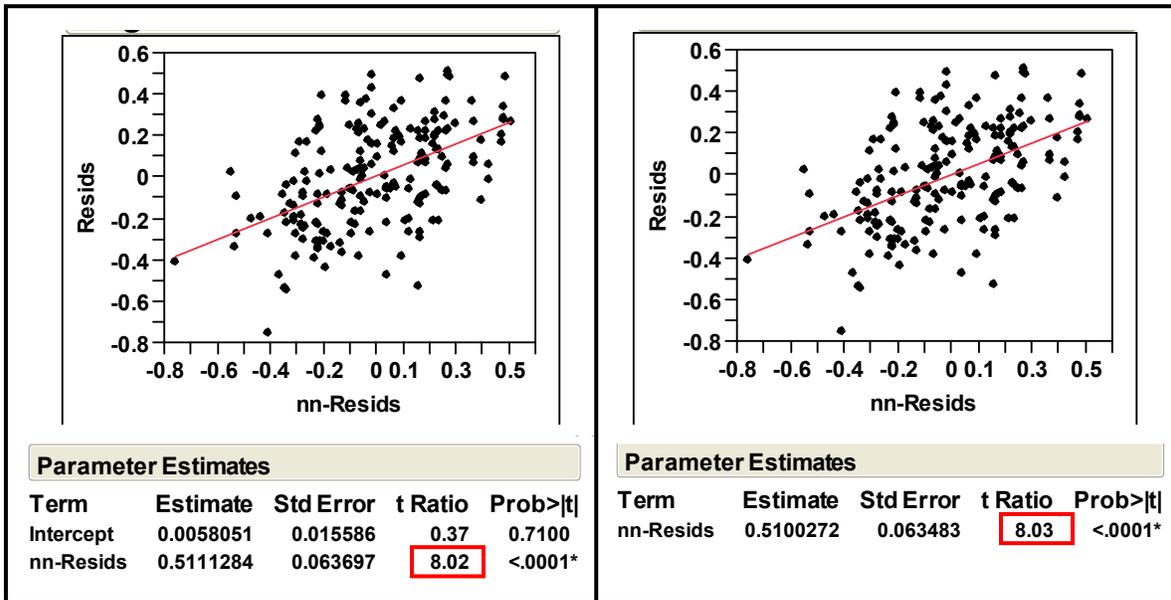
Figure 4.1. Regression with Intercept   **Figure 4.2. Regression with No Intercept**

---

[1] Recall from the reduced form of (4.5) that $u = (I_n - \rho W)^{-1} \varepsilon$ so that $E(u) = (I_n - \rho W)^{-1} E(\varepsilon) = 0$.

[2] This is established in expression (4.2.9) below.

However, it is also important to notice that from a regression viewpoint, models like (4.1.7) are seriously flawed. In particular, since the same random vector, u, appears both on the left and right hand sides, this regression suffers from what is called an "endogeneity problem". Here it can be shown that while $\hat{\rho}_W$ is still a consistent estimator of ρ, it is highly *inefficient* (i.e., tends to have a large variance). Nevertheless, we have already seen that the *p-values* derived from this regression are generally quite reasonable. So even though we will develop a more satisfactory Monte Carlo approach using this $\hat{\rho}_W$ statistic, the regression approach in (4.1.7) is generally quite robust and easy to perform.

4.1.2 Correlation Statistic

But given the inefficiency of $\hat{\rho}_W$ as an estimator of ρ, it is appropriate to seek alternative approaches to identifying ρ. One simple observation is to note in (4.1.5) that if $\rho > 0$ then for each $i = 1,..,n$, the ith row, $(Wu)_i$, of Wu contributes *positively* to the ith component, u_i, of u. So in this case, the random variables, u_i and $(Wu)_i$, should exhibit some degree of *positive correlation* for each $i = 1,..,n$. Similarly, u_i and $(Wu)_i$ should be *negatively correlated* for all i when $\rho < 0$. Hence it stands to reason that u_i and $(Wu)_i$ should also be *uncorrelated* for all i when $\rho = 0$. The argument here is slightly less obvious, but can be seen as follows. Since $\rho = 0$ implies that $u = \varepsilon$, we must also have $Wu = W\varepsilon$. But since the components of ε are all independent, and since $w_{ii} = 0$ implies that $(W\varepsilon)_i = \sum_{j \neq i} w_{ij}\varepsilon_j$ does *not* involve ε_i, it then follows that ε_i and $(W\varepsilon)_i$ must be *independent* for all i, and hence uncorrelated. In short, all pairs of random variables $[u_i,(Wu)_i]$ are *correlated with the same sign* (positive, negative, or zero).

Hence if the OLS regression residuals, $\hat{u}$, are taken to be a sample of u (so that $W\hat{u}$ is a sample of Wu) then all sample pairs $[\hat{u}_i,(W\hat{u})_i]$ must be correlated with the same sign. This suggests that, as a summary measure, the *sample correlation* between vectors, $\hat{u}$ and $W\hat{u}$, should reflect this common sign. Since all these random variables have zero means by construction,[3] we start by observing that the correlation between any *zero-mean* random variables, X and Y, is given by[4]

$$(4.1.9) \qquad \rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{E(XY)}{\sqrt{E(X^2)}\sqrt{E(Y^2)}}$$

Hence the appropriate sample estimator of $\rho(X,Y)$ is constructed as follows. If for any random samples, (x_i,y_i), $i = 1,..,n$ of (X,Y) we employ the natural sample estimators,

---

[3] Again, $(Wu)_i = \sum_j w_{ij}u_j$ implies from (4.1.8) that $E[(Wu)_i] = \sum_j w_{ij}E(u_j) = 0$.

[4] Recall that $\text{cov}(X,Y) = E(XY) - E(X)E(Y) = E(XY)$, so that $\text{var}(X) = \text{cov}(X,X) = E(X^2)$.

(4.1.10) $\hat{E}(XY) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\hat{E}(X^2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, $\hat{E}(Y^2) = \frac{1}{n}\sum_{i=1}^{n} y_i^2$

then the corresponding "plug in" estimator for $\rho(X,Y)$ is given by

(4.1.11) $\hat{\rho}(X,Y) = \dfrac{\hat{E}(XY)}{\sqrt{\hat{E}(X^2)}\sqrt{\hat{E}(X^2)}} = \dfrac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\ \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}}$

$$= \dfrac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\ \sqrt{\sum_{i=1}^{n} y_i^2}}$$

If we let $x = (x_1,..,x_n)'$ and $y = (y_1,..,y_n)'$, then in more common terminology, this estimator is designated simply as the *sample correlation, $r(x,y)$*, between x and y, i.e.,

(4.1.12) $r(x,y) = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\ \sqrt{\sum_{i=1}^{n} y_i^2}} = \dfrac{x'y}{\sqrt{x'x}\sqrt{y'y}} = \dfrac{x'y}{\|x\|\|y\|}$

In these terms, our second test statistic is given by the *sample correlation* between $\hat{u}$ and $W\hat{u}$, i.e.,

(4.1.13) $\boxed{r_W = r(\hat{u},W\hat{u}) = \dfrac{\hat{u}'W\hat{u}}{\|\hat{u}\|\|W\hat{u}\|}}$

We designate this as the *correlation statistic*, or more simply as *corr*.

4.1.3 Moran Statistic

Up to this point we have focused mainly on constructing statistics for estimating the value (or at least the sign) of ρ in model (4.5). But we have given little attention to how these statistics behave under the *null hypothesis*, H_0, in (4.6). One might suspect from the inefficiency of $\hat{\rho}_W$ that this statistic exhibits little in the way of "optimal" behavior under H_0. The sample correlation, r_W, does somewhat better in this respect. But from a statistical viewpoint, it suffers from a form of "incompatibility". For while the classical sample correlation statistic assumes that (x_i, y_i), $i = 1,..,n$ are independent random samples from the *same* statistical population (X,Y), this is not true of the samples $[\hat{u}_i,(W\hat{u})_i]$, $i = 1,..,n$. Even under the null hypothesis, where (4.7) implies that $(\hat{u}_i \equiv \hat{\varepsilon}_i : i = 1,..,n)$, are independently and identically distributed, this is not true of the samples, $(W\hat{\varepsilon})_i$, $i = 1,..,n$, which are neither independent nor identically distributed. So

there remains a question as to how well either of these statistics behaves with respect to *testing* H_0. In this context, we now introduce our final test statistic,

$$(4.1.14) \qquad \boxed{I_W = I(\hat{u}, W\hat{u}) = \frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}} = \frac{\hat{u}'W\hat{u}}{\|\hat{u}\|^2}}$$

designated as the *Moran statistic*, or more simply as *Moran's I*.[5] Here it is important to emphasize that expression (4.1.14) is different from the version of Moran's *I* in [BG, p.270] (also used in ARCMAP), which is designed for detecting autocorrelation in *Y* itself. This can in fact be viewed as the special case of (4.1) in which there only an "intercept" term with coefficient, μ, representing the common mean of the *Y* components, i.e.,

$$(4.1.15) \qquad Y = \mu 1_n + u$$

If *u* is again assumed to satisfy (4.5), then under the null hypothesis, $\rho = 0$, the "OLS" estimate in (4.3) reduces to the *sample mean* of *y*, i.e.,

$$(4.1.16) \qquad \hat{\mu} = (1_n'1_n)^{-1}1_n'y = (n)^{-1}\sum_{i=1}^{n} y_i = \bar{y}$$

Thus the "residuals" in (4.4) are here seen to be simply the deviations of the *y* components about this sample mean, i.e.,

$$(4.1.17) \qquad \hat{u} = y - \bar{y}1_n$$

So the appropriate version of Moran's *I* in this special case is seen to have the form,

$$(4.1.18) \qquad I_W = \frac{(y - \bar{y}1_n)'W(y - \bar{y}1_n)}{(y - \bar{y}1_n)'(y - \bar{y}1_n)} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

which is essentially the version used in [BG, p.270], except for the normalizing constant

$$(4.1.19) \qquad \alpha_W = \frac{n}{\sum_{i \neq j} w_{ij}}$$

For simplicity we have simply dropped this constant [as for example in Tiefelsdorf (2000,p.48)].[6]

---

[5] Be careful not to confuse this use of "*I*" with the *n*-square *identity matrix*, I_n.

[6] Notice that for the common case of *row normalized W* (with zero diagonal) it must be true that $\Sigma_{i \neq j} w_{ij} = \Sigma_{i=1}^{n}(\Sigma_{j \neq i} w_{ij}) = \Sigma_{i=1}^{n}(1) = n$, so this constant is unity.

While this statistic is more difficult to motivate by simple arguments,[7] it turns out to exhibit better statistical behavior with respect to testing H_0 then either of the statistics above.

4.1.4 Comparison of Statistics

By writing these three statistics side by side as

(4.1.20) $\hat{\rho}_W \; = \; \dfrac{\hat{u}'W\hat{u}}{\| W\hat{u} \|^2} \quad , \quad r_W \; = \; \dfrac{\hat{u}'W\hat{u}}{\| \hat{u} \| \, \| W\hat{u} \|} \quad , \quad I_W \; = \; \dfrac{\hat{u}'W\hat{u}}{\| \hat{u} \|^2}$

we see that they exhibit striking similarities. Indeed, since the numerators are identical, and since all denominators are *positive*, it follows that these three statistics *must always have the same sign*. Hence the differences between them are not at all obvious. But for testing purposes, the key issue is their relative behavior under the null hypothesis, H_0. To study this behavior, it is necessary to express $\hat{u}$ more explicitly as a random vector. To do so, observe first from (4.3) and (4.4) that

(4.1.21) $\hat{u} = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = [I_n - X(X'X)^{-1}X']Y$

$= [I_n - X(X'X)^{-1}X'](X\beta + u)$

$= (X\beta - X\beta) + [I_n - X(X'X)^{-1}X']u$

$= [I_n - X(X'X)^{-1}X']u$

$\Rightarrow \; \hat{u} = Mu$

where the matrix,

(4.1.22) $M \; = \; I_n - X(X'X)^{-1}X'$

is seen to be *symmetric*, i.e., $M = M'$. Notice also that

(4.1.23) $MM = [I_n - X(X'X)^{-1}X'][I_n - X(X'X)^{-1}X']$

$= [I_n - X(X'X)^{-1}X'] - X(X'X)^{-1}X' + X(X'X)^{-1}X'$

$= I_n - X(X'X)^{-1}X'$

$\Rightarrow \; MM = M$

---

[7] However, a compelling motivation of this statistic can be given in terms of the "concentrated likelihood function" used in *maximum likelihood estimation* of ρ. We shall return to this question in Section (??) after maximum likelihood estimation has been introduced.

Finally, to study the relative behavior of these estimators under H_0, recall from (4.5) that $\rho = 0$ implies $u = \varepsilon \sim N(0, \sigma^2 I_n)$, so that $\hat{u}$ takes now takes the form

$$(4.1.24) \qquad \hat{u} = M\varepsilon$$

with M given by (4.1.22), and satisfying $M' = M = MM$.[8] The estimators in (4.1.20) can then be expressed explicitly in terms of $\varepsilon \sim N(0, \sigma^2 I_n)$ as follows,

$$(4.1.25) \qquad \hat{\rho}_W = \frac{\varepsilon' MWM\varepsilon}{\|WM\varepsilon\|^2} \quad , \quad r_W = \frac{\varepsilon' MWM\varepsilon}{\|M\varepsilon\|\,\|WM\varepsilon\|} \quad , \quad I_W = \frac{\varepsilon' MWM\varepsilon}{\|M\varepsilon\|^2}$$

In terms of these specific representations under H_0 , the superiority of I_W relative to $\hat{\rho}_W$ and r_W can be illustrated in terms of the 190 Health Districts in the *English Mortality example* above. Here we choose W to be a *row normalized* weight matrix [expression (2.1.25) of Section 2 above] consisting of the *five nearest neighbors* of each district (with respect to centroid distance).[9] While the exact distributions of these statistics are difficult to obtain,[10] they can easily be approximated by simulating many random samples of ε . In Figure 4.3 below, the approximate sampling distributions of these three statistics are plotted using 10,000 simulated samples of ε with $\sigma^2 = 1$.[11]
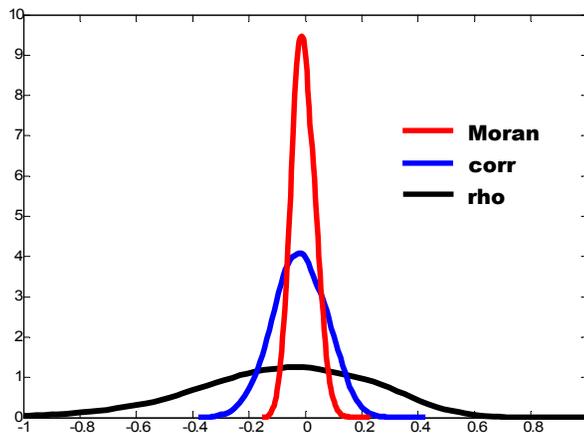
Figure 4.3. Sampling Distributions

Note first that while all three distributions are roughly centered on the true value, $\rho = 0$, there is actually some degree of bias in all three. The simulated means for these three

---

[8] The conditions $M' = M = MM$ together imply that M is an *orthogonal projection* matrix.

[9] Since the row sums are always 5, i.e., since $W1_n = (5)1_n$, it turns out that $\lambda_W = 5$ and thus that the *max-eigenvalue normalization* [(2.1.25) of Section 2 above] and *row normalization* for this particular W matrix are the same.

[10] Exact distribution results for Moran's *I* have been obtained by Tiefelsdorf (2000, Chap.7).

[11] Density estimation was done using the kernel-smoothing procedure, **ksdensity.m**, in MATLAB.

statistics are displayed in Table 4.1 below, and show that the mean of Moran's *I* is in fact an order of magnitude closer to zero than the other two. Moreover, the exact theoretical mean for Moran in this case [expression (4.1.29) below] can be calculated to be -0.00655, which shows that for a sample of size 10,000 these simulated values are quite accurate.

Moran	-0.0067
corr	-0.0198
rho	-0.0851

Table 4.1. Mean Values

But Figure 4.3 also suggests that relative bias among these three estimators is far less important that their *relative variances*. Indeed it is here that the real superiority of I_W is evident. While the variance of I_W under H_0 is known [see expression (4.1.31) below], its exact relation to the variances of r_W and $\hat{\rho}_W$ under H_0 is difficult to obtain analytically. But simulations with many examples show that these English Mortality results are quite typical. In fact, even for individual realizations of ε it is generally true that

(4.1.26) $|I_W(\varepsilon)| < |r_W(\varepsilon)| < |\hat{\rho}_W(\varepsilon)|$

While counterexamples show that (4.1.26) need not hold in all cases, this ordering was exhibited by *all* 10,000 simulations in the English Mortality example.

In summary, this example shows why Moran's *I* is by far the most widely used statistic for testing spatial autocorrelation. Given its relative unbiasedness and efficiency (small variance) properties under H_0, Moran's *I* tends to be the most reliable tool for detecting spatial autocorrelation.[12]

4.2. Asymptotic Moran Tests of Spatial Autocorrelation

Given the superiority of Moran's *I*, the most common procedure for testing H_0 is to use the *asymptotic normality* of I_W under H_0, first established by Cliff and Ord (1973) [see also Cliff and Ord (1981, pp.47-51), [BG], p.281 and Cressie (1993), p.442]. Since an asymptotic testing procedure using Moran's *I* is available in ARCMAP, it is of interest to develop this procedure here. But before doing so, it must be emphasized that the test used in ARCMAP is based on the version of Moran's *I* in expressions (4.1.18) and (4.1.19) above, which we here denote by

$$(4.2.1) \qquad \tilde{I}_W = \left(\frac{n}{\sum_{i \neq j} w_{ij}} \right) \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

---

[12] For a deeper discussion of its optimality properties, see Section 4.3.2 in Tiefelsdorf (2000).

The mean and variance of this statistic under H_0 are given in [BG, p.281]. For our present purposes, it is enough to observe that the mean of $\tilde{I}_W$ has a very simple form,

$$(4.2.2) \qquad E(\tilde{I}_W) = -\frac{1}{n-1}$$

which is made possible precisely by introducing the normalizing factor, α_W, in (4.1.19). This yields an expression which is *independent* of W (and thus serves to motivate the use of α_W). Note also that like the English Mortality example in Table 4.1 above, this result shows there is *always* a slight negative bias in $\tilde{I}_W$ under H_0, which shrinks to zero as n becomes large.

However, the simplicity of (4.1.2) disappears when analyzing I_W for *regression residuals*, $\hat{u} = M\varepsilon$, under H_0. Here even the mean of I_W has the more complex form:[13]

$$(4.2.3) \qquad E(I_W) = \tfrac{1}{n-k} tr(MW)$$

where M is given in terms of the $n \times k$ data matrix, X, by (4.1.22), and where the *trace*, $tr(A)$, of any n-square matrix, $A = (a_{ij})$, is defined to be the sum of its diagonal elements, i.e.,

$$(4.2.4) \qquad tr(A) = \sum_{i=1}^{n} a_{ii}$$

So in the case of regression, $E(I_W)$, is seen to depend not only on W but also on the particular data matrix, X. This is also true for the variance of I_W, which has the more complex form:

$$(4.2.5) \qquad var(I_W) = \frac{tr(MWMW') + tr(MWMW) + [tr(MW)]^2}{(n-k)(n-k+2)} - [E(I_W)]^2$$

So before considering the testing procedure in ARCMAP, it is appropriate to consider the more general asymptotic test for regression residuals.

---

[13] The mean (4.2.3) and variance (4.2.5) of I_W under H_0 are taken from Tiefelsdorf (2000, p.48). The original derivations of these moments (using the normalizing factor, α_W) can be found in Cliff and Ord (1981, Sections 8.3.1 and 8.3.2).

_____

4.2.1 Asymptotic Moran Test for Regression Residuals

Given the above mentioned asymptotic normality property of I_W under H_0, it follows that if expressions (4.2.3) and (4.2.5) are now employed to standardize this statistic as,

$$(4.2.6) \qquad Z_W = \frac{I_W - E(I_W)}{\sqrt{\mathrm{var}(I_W)}}$$

then under "appropriate conditions", this standardized statistic, Z_W, should be approximately standard normally distributed, i.e.,

$$(4.2.7) \qquad Z_W \approx_d N(0,1)$$

where (as in Section 3.1.3 of Part II) the notation, $\approx_d$ means "is approximately distributed as". A more detailed description of these "appropriate conditions" will be given at the end of Section 4.2.2 below. So for the present, we simply assume that the approximation in (4.2.7) is valid.

Given this assumption, the appropriate testing procedure is operationalized in the MATLAB program, **moran_test_asymp.m**. To apply this program to the English Mortality example, let **y** = "lnMI" and **x** = "lnJarman" (as in Section 1.3 above) and denote the 5-nearest neighbor weight matrix by **W**. This test can then be run with the command:

>> moran_test_asymp(y,x,W);

Note that this program actually runs the OLS regression, calculates the residuals, $\hat{u}$, and then calculates Z_W in (4.2.6). The test results are reported as screen outputs:

Moran = 0.46405

Zval = 11.3044

Pval < 10^(-323)

Here the calculated value, Z_W, is denoted by **Zval** and is seen to be more than 11 standard deviations above the mean. This suggests that there is simply no chance at all that these residual values (shown in Figure 4.4 below) could be spatially independent.[14]

4.2.2 Asymptotic Moran Test in ARCMAP

As mentioned above, the Moran test used in ARCMAP relies on $\tilde{I}_W$ in (4.2.1) rather than I_W, and essentially tests whether a given set of spatial data, y, can be distinguished from independent normal samples. This procedure can again be illustrated using the English

_____

[14] In fact, the p-value here is so small that it is reported as "0" in MATLAB. In such cases, the program simply reports "**Pval < 10^(-323)**", which is roughly the smallest number treated as nonzero in MATLAB.

_____

Mortality data. But in doing so, it must be born in mind that these regression residuals, $\hat{u}$, are now treated as the basic data set "y" itself. This is always possible in the case of OLS residuals since by definition the "sample mean", $\frac{1}{n}1'_n\hat{u}$, of such residuals is *identically zero*. This is a consequence of the following property of the projection matrix, M,

(4.2.8) $M X = [I_n - X(X'X)^{-1}X']X = X - X = O_n$

which together with the definition, $X = (1_n, x_1, .., x_k)$, implies in particular that $M1_n = 0$. But for any realized value, y, of Y it then follows from (4.1.21) that

(4.2.9) $\hat{u} = My \implies 1'_n\hat{u} = 1'_n My = (M1_n)'y = 0$

and thus that $\frac{1}{n}1'_n\hat{u}$ is always zero. This means that we can set $\hat{u} = y$ in (4.1.17) then $\overline{y} = 0$ and obtain no immediate contradictions. However, it is important to emphasize that the mean and covariance of $\tilde{I}_W$ are in principle very different from those of I_W, depending on the explanatory data, X.

Given this observation, we now proceed to test for spatial autocorrelation in $\hat{u}$ by using $\tilde{I}_W$ rather than I_W. To do so, the OLS residuals of the regression of Myocardial Infarction rates on the Jarman Index must first be imported to ARCMAP and joined to the **Eng_Mort.shp** file as a new column, say **resid**, and saved as a new shapefile, say **OLS_resids**. These residuals are shown in Figure 4.4 below, with positive residuals in *red*, negative in *blue*, and with all values "close to zero" (i.e., within half a standard deviation) shown in *white*. Here it is clear that while the Jarman Index is certainly a significant predictor of Myocardial Infarction, these unexplained residuals are highly correlated in space. Recall from the simple nearest-neighbor test that this correlation was more significant than the Jarman Index itself. We now show that this degree of significance is in fact even greater than in that simple heuristic test.
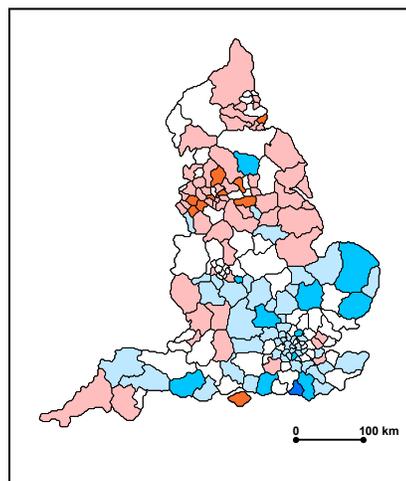
Figure 4.4. OLS Residuals

For this illustration, we again use the above weight matrix, *W*, consisting of the first 5 nearest neighbors of each district. To use this weight matrix in ARCMAP, it must first be converted to a *text file* using the MATLAB program, **arcmap_wt_matrix.m**, with the command:

>> L = arcmap_wt_matrix(W);

Here you must be sure that **W** is not in "sparse" form (which can be seen by displaying the first row of **W**). If it is, then use the command, **W = full(W),** to convert it to a full matrix. The matrix output, **L**, should then have initial rows of the form:

1	**2**	**0.2**
1	**3**	**0.2**
1	**8**	**0.2**
1	**15**	**0.2**
1	**16**	**0.2**
⋮	⋮	⋮

This shows in particular that the first 5 nearest neighbors of district 1 are districts (2,3,8,15,16). To import this matrix to ARCMAP, first open it in EXCEL and "clean" the numerical format to look like the above. ARCMAP also requires an ID for these values, which can be accomplished in three steps:

(i)   First add a new column to the attribute table (say next to **resid**) labeled **ID** and use the calculator (with "short integer") to create values (1 2 3 …) by setting **ID = [FID] + 1**.

(ii)   Now add a new row at the top of the matrix, **L**, in EXCEL, and write the identifier name, ID, so that **L** is now of the form:

ID		
1	2	0.2
1	3	0.2
1	8	0.2
⋮	⋮	⋮

(iii)   Finally, save this as a text file, say **Wnn_5.txt**, (to indicate that it includes the first 5 nearest neighbors). This file will be used by ARCMAP below.

To apply the Moran test to the OLS residuals, **resid**, in ARCMAP, follow the path:

> **ArcToolbox > Spatial Statistics Tools**
> **> Analyzing Patterns**
> **> Spatial Autocorrelation (Morans I)**

In the "Spatial Autocorrelation" window that opens, fill in the shapefile and field, and be sure to check the "Generate Report" box. You are now going to use the option,

      "GET_SPATIAL_WEIGHTS_FROM_FILE"

in the "Conceptualization of Spatial Relationships" window. (It is this option which makes the ARCMAP test worthwhile!). Here, browse to the text file, **Wnn_5.txt**, constructed above. The relevant portion of the Spatial Autocorrelation window should now look as shown below (where the last file path will of course vary):
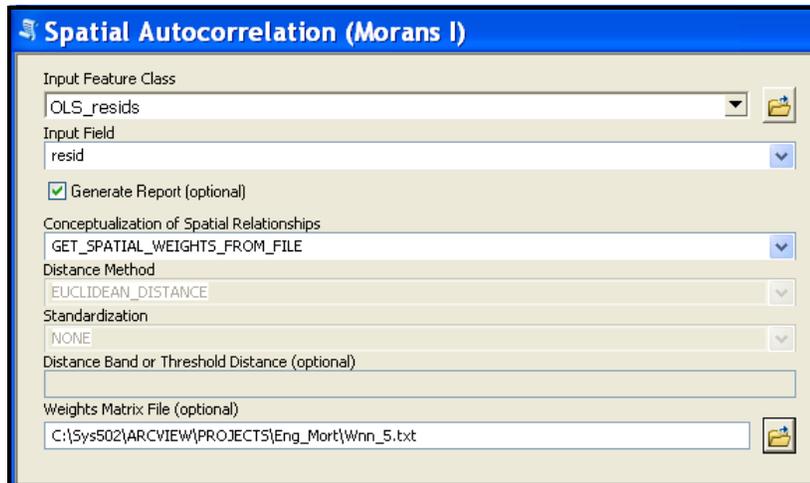
Figure 4.5. Spatial Autocorrelation Window

Click **OK**, and when the procedure terminates, you will get a report displayed. The most relevant portion of this report is shown in Figure 4.7 below:
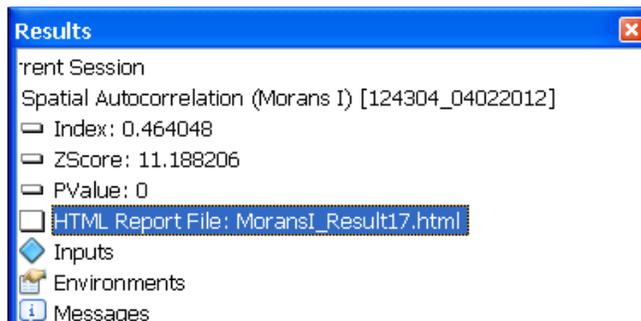
Figure 4.7. Moran Test Report

Before proceeding further, notice that while the value of Moran's *I*, (**Index = 0.464048**), is the same as in Section 4.2.1 above, the Z value **(ZScore = 11.188206)** is slightly different. This is because the mean and variance used to standard Moran's *I* are *different*

in these two tests. Rather than using (4.2.3) and (4.2.5), the values used are those in [BG, p.281], including (4.2.2) for the mean. In the present case, spatial autocorrelation is so strong that there is little difference between these results. But this need not always be the case.

It is also important to note that while this report contains all key test information, there is a much better graphical representation that can be obtained by clicking the "HTML Report File" that is shown here in blue. This graphic is shown in Figure 4.9 below.
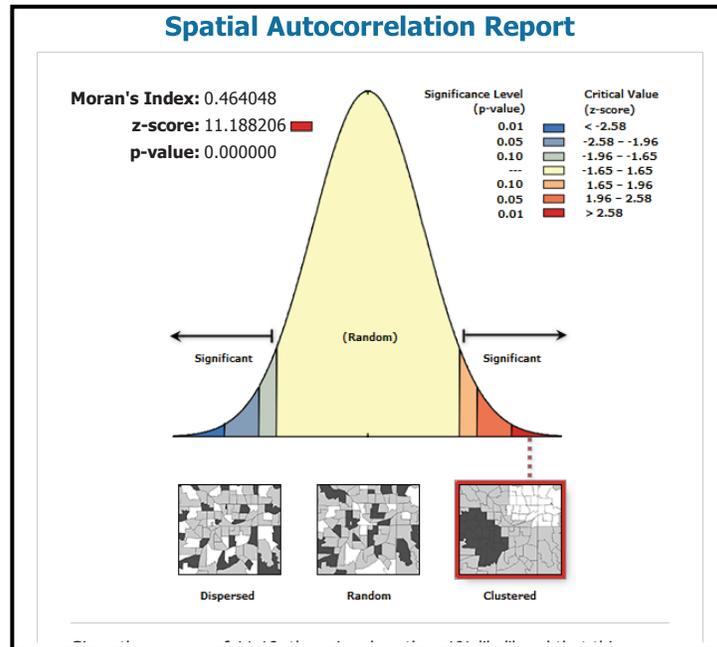
Figure 4.9. Moran Test Report

This graphic facilitates the interpretation of the results by making it abundantly clear (in the present case) that these test results show positive spatial correlation that is even more significant than that of the heuristic nearest-neighbor approach used previously.

But it should be emphasized that while spatial correlation is visually evident in Figure 4.5 above, this will not always be the case. Moreover, it should also be stressed that the Moran statistics, I_W and $\tilde{I}_W$ (as well as $\hat{\rho}_W$ and r_W) are defined only with respect to a *given* weight matrix, W. Hence it is advisable to use a number of *alternative* weight matrices when testing for spatial autocorrelation. For example, one might try alternative numbers of neighbors (say 4, 5, and 6), or more generally, weight matrices involving both *distance-based* and *boundary-based* notions of spatial proximity. A general rule-of-thumb is to try three substantially difference matrices, (W_1, W_2, W_3), that cover a range of potentially relevant types of proximity. If the results for all three matrices are comparable (as will surely be the case in the English Mortality example), then this will help to substantiate these results. On the other hand, if there are significant differences in these

_____

results, then an analysis of these differences may serve to yield further information about the underlying structure of the unobserved spatial dependencies.

Finally it should be emphasized that, as with all asymptotic tests, these *asymptotic Moran tests* require that the number of samples (areal units) be "sufficiently large". Moreover, it is also required that the *W* matrix be "sufficiently sparse" (i.e., consist mostly of zero-valued cells) to ensure that the *Central Limit Theorem* is working properly. The present case with $n = 190$ spatial units, and with each row of *W* containing only 5 nonzero entries, this should be a reasonable assumption. But as with the Clark-Evans tests for random point patterns, it is often difficult to know how well this normal approximation is working.[15]

4.3. A Random Permutation Test of Spatial Autocorrelation

With this in mind, we now develop an alternative testing procedure based on Monte Carlo methods that is more computationally intensive, but requires essentially *no* assumptions about the distribution of test statistics under H_0. The basic idea is very similar to the "random relabeling" test of independence for point patterns in Section 5.6 of Part I. There we approximated the hypothesis of statistical independence by "spatial indistinguishability". Here we adopt the same approach by postulating that if the particular spatial arrangement of sample points doesn't matter, then neither should the labeling of these points. More specifically, in the vector of regression residuals, $\hat{u} = (\hat{u}_1,..,\hat{u}_n)$, it shouldn't matter which residual is labeled as "$\hat{u}_1$", "$\hat{u}_2$", or "$\hat{u}_n$". If so, then regardless of what the joint distribution of these residuals $(\hat{u}_1,..,\hat{u}_n)$ actually is, each relabeling $(\hat{u}_{\pi_1},..,\hat{u}_{\pi_n})$ of these residuals should constitute an equally likely sample from this distribution. So under this *spatial invariance hypothesis*, H_{SI}, we may generate the sampling distribution for any statistic, say $S(\hat{u}_1,..,\hat{u}_n)$, under H_{SI} by simply evaluating $S(\hat{u}_{\pi_1},..,\hat{u}_{\pi_n})$ for many random relabelings, π, of $(\hat{u}_1,..,\hat{u}_n)$. As we have seen for point-pattern tests, this hypothesis can then be rejected if the observed value, $S(\hat{u}_1,..,\hat{u}_n)$, appears to be an unusually high or low value from this sampling distribution.

Before operationalizing this procedure, it is important to stress that it is applicable to any statistic constructible from this residual data. Hence, in the same way that different weight matrices, *W*, can be used to reflect alternative notions of spatial proximity, it is advisable to use a range of alternative test statistics for H_{SI}. Indeed, this is precisely why the *rho statistic* and *correlation statistic* were developed above. While Moran's *I* appears to be the best choice when residuals are multi-normally distributed, this is less clear in the present nonparametric setting. So it seems reasonable to check the results for I_W with those of r_W and $\hat{\rho}_W$. However, it should also be emphasized that a substantial body of simulation results in the literature suggest that Moran's *I* tends to be robust with respect

_____

[15] For further discussion of these issues, see for example Tiefelsdorf (2000, Section 9.4.1), Anselin and Rey (1991) and Anselin and Florax (1995).

_____

to violations of normality.[16] So while we shall report results for all three statistics, Moran's *I* tends to be the most reliable of these three.

4.3.1 SAC-Perm Test

With this overview, we now outline the steps of testing procedure for H_{SI}, designated as the *permutation test of spatial autocorrelation*, or more simply the *sac-perm test*. For convenience we maintain the general notation, S, which can stand for either I, $\hat{\rho}$, or r. Since higher positive values of each of these three statistics correspond to higher levels of *positive* spatial autocorrelation, we assume that *S* exhibits this same ordering property. In this setting, our test is designed as a *one-tailed test of positive spatial autocorrelation* (paralleling the one-tailed test of clustering for K-functions). In particular, significant *positive* (*negative*) spatial autocorrelation will again be reflected by *low* (*high*) p-values. As with the asymptotic Moran tests above, this sac-perm test is defined with respect to a *given* spatial weight matrix, *W*. Finally, it should be noted that while the notation, u, will be used to represent the given residual data in this procedure, virtually all applications will be in terms of *OLS residuals*, i.e., $u = \hat{u}$. With these preliminary observations, the steps of this testing procedure are as follows:

Step 1. Let $u^0 = (u_1,..,u_n)$ denote the vector of *observed residuals*, and construct the corresponding value, $S^0 = S(u^0)$, of statistic, S.

Step 2. Simulate N random permutations, $\pi^j = (\pi_1^j,..,\pi_n^j)$, of the integers $(1,..,n)$.[17]

Step 3. For each permutation, π^j, construct the corresponding permuted data vector, $u(\pi^j) = (u_{\pi_1^j},..,u_{\pi_n^j})$, and the resulting value of S, denoted by $S^j = S[u(\pi^j)]$, $j = 1,..,N$.

Step 4. Rank the values $(S^0,S^1,..,S^N)$ from high to low, so that if S^j is the kth highest value then $rank(S^j) = k$.

Step 5. If $rank(S^0) = k$ then define the *p-value* for this test to be

(4.3.1) $\alpha = k/(N+1)$

   (i)   If α is *low* (say $\alpha \leq 0.05$) then conclude that there is significantly *positive* spatial autocorrelation at the α -*level of significance*.

---

[16] See for example the Monte Carlo results in Anselin and Rey (1991) and Anselin and Florax (1995).

[17] So if *n* = 3 then the first permutation of (1,2,3) might be $\pi^1 = (\pi_1^1, \pi_2^1, \pi_3^1) = (2,3,1)$.

(ii)   Conversely, if α is *high* (say $\alpha \geq 0.95$) then conclude that there is significantly *negative* spatial autocorrelation at the $(1-\alpha)$-*level of significance.*

(iii)   If neither (i) or (ii) hold, then conclude that the spatial independence hypothesis, H_{SI}, *cannot be rejected.*

The "cutoff" levels for significantly positive or negative spatial autocorrelation are intentionally left rather vague. Indeed the entire nature of this sac-perm test is meant to be exploratory in nature.

4.3.2 Application to the England Mortality Data

Recall from the asymptotic Moran tests in Section 4.2 that there was extremely strong autocorrelation in the OLS residuals of the *England Myocardial Infarction* data when regressed on the *Jarman Index*. We now reproduce those results in MATLAB using **sac_perm**. Here the data can be found in the workspace, **eng_mort.mat**, where the OLS residuals are in the vector, **Res**. Finally, recall that desired weight matrix, **Wnn_5**, was already constructed for Eire in Section 2.2.1 above. So the only difference here is that **L** is now a 190x2 matrix of coordinates for English Health Districts. To construct a *sac-perm test* of the residual data, **Res**, using this weight matrix, we can employ 9999 random permutations with the command:

>> **sac_perm(Res, Wnn_5, 9999);**

The screen output of this program is shown below:

```
RANGE OF RANDOM-PERMUTATION INDEX VALUES:

INDEX        Moran           corr           rho
MAX          0.1939        0.3325         0.6799
MIN         -0.1444       -0.3797        -0.9987


TABLE OF SIGNIFICANCE LEVELS:

INDEX         VALUE         SIGNIF
Moran         0.4640        0.0001
corr          0.6324        0.0001
rho           0.8618        0.0001
```

Here the key outputs are the *significance levels* for the three test statistics (**Moran**, **corr**, **rho**). Notice that (as expected), these values are each *maximally significant*, i.e., they are higher than the values for *all* of the 9999 random permutations simulated. In fact, they are *much* higher, as can be seen by comparing them with the range of values displayed above. For example, the important Moran value, **0.4640**, is seen to be well above the range of values, **- 0.1444** to **0.1939**, reported for all 9999 permutations. Note also that the

ranges of **corr** and **rho** are successively larger than this, in a manner consistent with expression (4.1.26) for the asymptotic Moran test.

As expected, the Moran value, **0.4640** , is the same as that for the asymptotic tests above, confirming that the same weight matrix and calculation procedure are being used. Moreover, the extremely significant p-value reported for those tests is consistent with the present fact that this Moran value is way above the simulated range. This shows that if the number of permutations were increased well beyond, 9999, the same maximally-significant results would almost surely persist.

Finally, just to show that normality of I_W persists under random permutations for samples this large, we have plotted the histogram for the 9999 simulated values of I_W (ranging from -0.1444 to 0.1939), together with the observed value, 0.4640, shown in red.
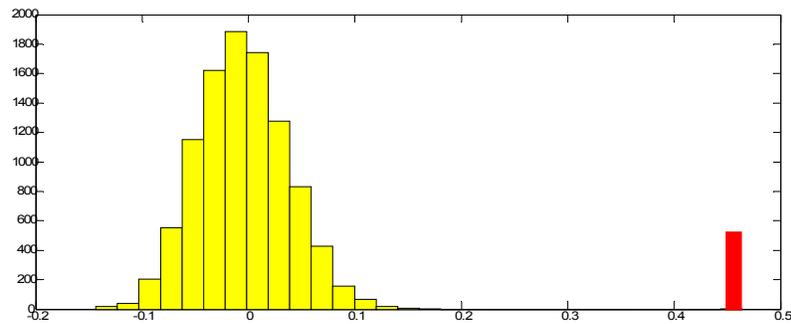
Figure 4.10. SAC-Perm Test for I_W

This plot also serves to further dramatize the significance of spatial autocorrelation for these particular regression residuals.