GEODA DIAGNOSTICS FOR SPATIAL REGRESSION

1. R-Square

Let the *residuals* of the spatial regression be denoted by

(1) $\hat{\varepsilon} = (\hat{\varepsilon}_1,..,\hat{\varepsilon}_n)' = y - \hat{y} = (y_1,..,y_n)' - (\hat{y}_1,..,\hat{y}_n)'$

and let the *error sum of squares* (*ESS*) and *total sum of squares* (*TSS*) be given by

(2) $ESS = \hat{\varepsilon}'\hat{\varepsilon}$

(3) $TSS = (y - \overline{y}1)'(y - \overline{y}1)$

then in GEODA,

(4) $R\text{-}square = 1 - \dfrac{ESS}{TSS}$ (= **0.7386** for Eire example)

But if the *model sum of squares* (*MSS*) is denoted by

(5) $MSS = (\hat{y} - \overline{\hat{y}}1)'(y - \overline{\hat{y}}1)$

then the fraction of *TSS* that is "explained" by the model is

(6) $R^2 = \dfrac{MSS}{TSS}$ (= **0.5733** for Eire example)

So which do we choose?

2. Akaike Information Criterion

If the *log likelihood* of the model estimates is denoted by

(7) $L = L(\hat{\theta} \mid y, X)$ (= **-48.0591** for Eire example)

where $\hat{\theta} = (\hat{\theta}_1,..,\hat{\theta}_k)$ is the vector of parameter estimates, then the *Akaike Information Criterion* (*AIC*) is given by

(8) $AIC = 2\cdot\{-L + k\}$ (= $2\cdot\{48.0951 + 3\} = 102.118$ for Eire example in GEODA)

[Note: In the present case, $\hat{\theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2, \hat{\lambda})$, so this *should* be $k = 4$]. Intuitively, *AIC* is an "adjusted log likelihood" that is penalized for the number of parameter fitted, in a manner analogous to *adjusted R-Square*.

3. Breusch-Pagan Test of Heteroscedasticity

The Breusch-Pagan Test considers heteroscedastic variance models of the form

(9) $$\sigma_i^2 = \sigma^2(\alpha_0 + \alpha' x_i) \;\Rightarrow\; \frac{\sigma_i^2}{\sigma^2} = \alpha_0 + \alpha' x_i$$

where $x_i = (1, x_{i1}, .., x_{ik})$ is the vector of explanatory variables (plus intercept) for observation i and $\sigma_i^2 = \text{var}(\varepsilon_i)$. The appropriate null hypothesis is then, $H_0 : \alpha = 0$. To test this hypothesis, observe that since $\text{var}(\varepsilon_i) = E(\varepsilon_i^2)$, the residual, $\hat{\varepsilon}_i^2$, constitutes a *one-sample estimate* of σ_i^2. If the *mean-square error* is denoted by

(9) $$s^2 = \tfrac{1}{n}\hat{\varepsilon}'\hat{\varepsilon} = \tfrac{1}{n} ESS$$

then under H_0, the sample vector

(10) $$r = \left(\frac{\hat{\varepsilon}_1^2}{s^2}, .., \frac{\hat{\varepsilon}_n^2}{s^2} \right)$$

is a natural estimate of $(\sigma_1^2/\sigma^2, .., \sigma_n^2/\sigma^2)$. Hence if one regresses r on the set of explanatory variables, $X = [1, x_1, .., x_k]$, then "significantly large" values for the *model sum of squares* (*MSS*) of this regression (under the null hypothesis H_0) indicate that model (9) fits better than would be expected under H_0. The appropriate *Breusch-Pagan statistic*, S_{BP}, is thus taken to be twice this *MSS*,

(11) $$S_{BP} = \tfrac{1}{2} MSS = \tfrac{1}{2}\left[r'X(X'X)^{-1}X'r - n \right]$$ (= 0.0743 for Eire example)

which can be shown to be asymptotically distributed χ_{k+1}^2 under H_0. In the Eire case, this value is not sufficiently high to suggest the presence of significant heteroscedasticity (*P-value* = .785)