SIMPLE EXAMPLE OF THE EFFECTS OF SPATIAL AUTOCORRELATION

Given the simple spatial model:

(1) $$Y(s) = \mu + \varepsilon(s) \ , \ s \in \{s_1,..,s_n\}$$

suppose it is assumed that $\varepsilon(s) \underset{iid}{\sim} N(0,\sigma^2)$ so that in matrix form we have:

(2) $$\boxed{Y = \mu \cdot 1 + \varepsilon \ , \quad \varepsilon \sim N(0,\sigma^2 I)}$$

[with $1 = (1,..,1)'$]. Then it is well known that the *sample-mean estimator*,

(3) $$\bar{Y} = \tfrac{1}{n}\sum_{i=1}^{n} Y_i$$

is the *minimum-variance linear unbiased estimator* for μ, and in particular, that this minimum variance is given by

(4) $$\boxed{\mathrm{var}(\bar{Y}) = \frac{\sigma^2}{n}}$$

But suppose that in reality there is *positive spatial autocorrelation* among the residuals in (2) so that in fact:

(5) $$\mathrm{cov}(\varepsilon) = \begin{bmatrix} \mathrm{cov}(\varepsilon_1,\varepsilon_1) \cdots \mathrm{cov}(\varepsilon_1,\varepsilon_n) \\ \vdots \quad \ddots \quad \vdots \\ \mathrm{cov}(\varepsilon_n,\varepsilon_1) \cdots \mathrm{cov}(\varepsilon_n,\varepsilon_n) \end{bmatrix} = \begin{pmatrix} \sigma^2 \cdots \sigma_{1n} \\ \vdots \ \ddots \ \vdots \\ \sigma_{n1} \cdots \sigma^2 \end{pmatrix} \geq \sigma^2 I$$

with $\sigma_{ij} > 0$ for many distinct (i,j) pairs. Then [as in expression (4.10.3) of the NOTEBOOK] it follows that since $\mathrm{cov}(Y) = \mathrm{cov}(\varepsilon)$, the true variance of $\bar{Y}$ is given by

(6) $$\mathrm{var}(\bar{Y}) = \mathrm{var}\left(\tfrac{1}{n}\sum_{i=1}^{n} Y_i\right) = \tfrac{1}{n^2}\left(\sum_{i=1}^{n} \mathrm{var}(Y_i) + \sum_i \sum_{j \neq i} \mathrm{cov}(Y_i,Y_j)\right)$$

$$= \tfrac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(Y_i) + \tfrac{1}{n^2}\sum_i \sum_{j\neq i} \mathrm{cov}(Y_i,Y_j) \ = \ \tfrac{1}{n^2}\sum_{i=1}^{n} \sigma^2 + \tfrac{1}{n^2}\sum_i \sum_{j \neq i} \sigma_{ij}$$

$$= \tfrac{1}{n^2}(n\sigma^2) + \tfrac{1}{n^2}\sum_i \sum_{j\neq i} \sigma_{ij} \ = \ \boxed{\frac{\sigma^2}{n} + \tfrac{1}{n^2}\sum_i \sum_{j \neq i} \sigma_{ij}}$$

which together with the positive spatial dependencies shows that:

$$(7) \qquad \boxed{\mathrm{var}(\bar{Y}) \;>>\; \frac{\sigma^2}{n} \;\Rightarrow\; \sigma(\bar{Y}) \;>>\; \frac{\sigma}{\sqrt{n}}}$$

Hence if we consider, say, a 95% confidence interval for the true mean, μ, then the *actual* interval is given by

$$(8) \qquad CI_{actual} = [\bar{Y} \pm (1.96)\,\sigma(\bar{Y})]$$

rather than the *assumed* interval

$$(9) \qquad CI_{assumed} = \left[\bar{Y} \pm (1.96)\tfrac{\sigma}{\sqrt{n}}\right]$$

So for any given estimate, $\bar{y}$, this implies from (7) that