MULTILEVEL SPATIAL MODELING

The following simple example of a multilevel model is taken from Bingenheimer and Raudenbush (2004). This example is a small version of ongoing studies on obesity in the U.S., and illustrates the use of multilevel models to capture effects of neighborhood attributes as well as individual attributes on obesity. Suppose that one wishes to study variations in body mass index (BMI) among young adults, focusing on sex differences and differences related to the presence or absence of fast-food restaurants in individuals' neighborhoods. Let Y_{ij} denote the BMI of individual i living in neighborhood j, and let the indicator variable, X_{ij}, be unity if i is male and zero if i is female. Then a simple linear model relating BMI to sex differences is given by:

(1) $$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij} \ , \ e_{ij} \underset{iid}{\sim} N(0,\sigma^2) \ , \ i = 1,..,n_j \ , \ j = 1,..,J$$

The subscript j on the intercept indicates that there is some type of "neighborhood" effect present. If the β_{0j} are constants for each j, then model (1) is called a *fixed neighborhood effects* model. If the β_{0j} are independent and identically distributed random variables, then model (1) is called a *random neighborhood effects* model. More generally, each β_{0j} may itself be determined by a linear model involving neighborhood attributes. In particular, if the indicator variable, W_j, is unity whenever a fast-food restaurant is present in j, and is zero otherwise, then one might postulate that

(2) $$\beta_{0j} = \alpha_0 + \alpha_1 W_j + u_j \ , \ u_j \underset{iid}{\sim} N(0,\tau^2) \ , \ j = 1,..,J$$

If in addition it is assumed that $\mathrm{cov}(u_j, e_{ij}) = 0,$ for all i and j, then model [(1),(2)] represents a simple example of a *multilevel* (*hierarchical*) *model*, involving two levels. By substituting (2) into (1) we obtain the single-equation model:

$$Y_{ij} = \alpha_0 + \alpha_1 W_j + \beta_1 X_{ij} + \varepsilon_{ij} \ ,$$

(3) $$\varepsilon_{ij} = u_j + e_{ij} \ , \ e_{ij} \underset{iid}{\sim} N(0,\sigma^2) \ , u_j \underset{iid}{\sim} N(0,\tau^2)$$

$$\mathrm{cov}(u_j, e_{ij}) = 0, \quad i = 1,..,n_j \ , \ j = 1,..,J$$

The first line is simply a standard linear model involving both individual and neighborhood attributes. Thus, the key feature of this model is the *error term*,

$\varepsilon_{ij} = u_j + e_{ij}$, which involves both individual and neighborhood effects. In particular, observe that since $E(\varepsilon_{ij}) = E(u_j) + E(e_{ij}) = 0$,

(4) $$\text{var}(\varepsilon_{ij}) = E(\varepsilon_{ij}^2) = E\left[(u_j + e_{ij})^2\right] = E(u_j^2) + 2E(u_j e_{ij}) + E(e_{ij}^2)$$

$$= E(u_j^2) + (0) + E(e_{ij}^2) = \tau^2 + \sigma^2$$

and that for all distinct individuals i and k in the same region, j,

(5) $$\text{cov}(\varepsilon_{ij}, \varepsilon_{kj}) = E\left[(u_j + e_{ij})(u_j + e_{kj})\right] = E\left[u_j^2 + u_j(e_{ij} + e_{kj}) + e_{ij}e_{kj}\right]$$

$$= E(u_j^2) + (0) + (0) = E(u_j^2) = \tau^2$$

Finally since $\text{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$ whenever $j \neq j'$, it follows that the *covariance matrix*, $V = \text{cov}\left[(\varepsilon_{ij})\right]$ has block-diagonal form:

(6) $$V = \begin{pmatrix} \tau^2+\sigma^2 & \cdots & \tau^2 & & & & \\ \vdots & \ddots & \vdots & & & & \\ \tau^2 & \cdots & \tau^2+\sigma^2 & & & & \\ & & & \ddots & & & \\ & & & & \tau^2+\sigma^2 & \cdots & \tau^2 \\ & & & & \vdots & \ddots & \vdots \\ & & & & \tau^2 & \cdots & \tau^2+\sigma^2 \end{pmatrix}$$

Hence this multilevel covariance structure represents an alternative way of capturing *unobserved spatial dependencies*. In this case, all dependencies are assumed to be *within-neighborhood* dependencies. The strength of this dependency is often represented in terms of the *within-neighborhood correlations*:

(7) $$\rho(\varepsilon_{ij}, \varepsilon_{kj}) = \frac{\text{cov}(\varepsilon_{ij}, \varepsilon_{kj})}{\sigma(\varepsilon_{ij})\sigma(\varepsilon_{kj})} = \frac{\tau^2}{\sqrt{(\tau^2+\sigma^2)} \cdot \sqrt{(\tau^2+\sigma^2)}} = \frac{\tau^2}{\tau^2+\sigma^2}$$

Reference:

Bingenheimer, J.B. and S. Raudenbush (2004) "Statistical and substantive inferences in public health: Issues in the application of multilevel models", *Annual Review of Public Health*, 25: 53-77.