Appendix: Multivariate Regression

The multivariate generalized least squares regression model is $\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, and defined in Section 2.2.1. Based on the analysis results of Section 3.2, we assume a zero mean, normally distributed error with variance-covariance structure given by $\mathrm{cov}(\mathbf{e}_{(i)}, \mathbf{e}_{(k)}) = \mathbf{\Sigma}_{e(i,k)} = \sigma_{ik}\mathbf{V}$ where $i, k = 1, \ldots, m$. This covariance structure is separable into two aspects. The quantity σ_{ik} is the overall magnitude of variances/covariances between the errors on the ith and kth response variables. The matrix $\mathbf{V}$ provides for the spatial correlation structure within and between response variables. Introduction of the correlation matrix $\mathbf{V}$ makes this a problem in multivariate *generalized* least-squares (MV GLS) regression.

The most straightforward solution to the MV GLS problem of estimating $\mathbf{b}$ uses variable transformation. First, find the square root matrix $\mathbf{K}$ such that $\mathbf{K}\mathbf{K} = \mathbf{V}$. Then transform the response and explanatory variables as:

$$\tilde{\mathbf{Y}} = \mathbf{K}^{-1}\mathbf{Y}, \qquad \tilde{\mathbf{X}} = \mathbf{K}^{-\mathbf{1}}\mathbf{X}, \qquad \tilde{\mathbf{e}} = \mathbf{K}^{-\mathbf{1}}\mathbf{e}.$$

This renders the problem one of multivariate *ordinary* least squares regression (MV OLS), such that $\tilde{\mathbf{e}}$ has an covariance structure following $\mathbf{\Sigma}_{e(i,k)} = \sigma_{e(i,k)}\mathbf{I}$. Standard multivariate regression results for MV OLS regression are available in any text in multivariate statistics (e.g., Johnson and Wichern 2001, ch 7).

The major regression results used in this study are the following. Regression coefficients are obtained as

$$\hat{\mathbf{b}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}.$$

The variance-covariance matrix of the regression coefficients is

$$\text{cov}(\hat{\mathbf{b}}_{(i)}, \hat{\mathbf{b}}_{(k)}) = \hat{\mathbf{\Sigma}}_{b(i,k)} = \hat{\sigma}_{e(i,k)}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$$

where $\hat{\sigma}_{ik}$ are the elements of the estimated variance-covariance matrix

$$\hat{\mathbf{\Sigma}}_e = \frac{1}{n-p}\hat{\tilde{\mathbf{e}}}'\hat{\tilde{\mathbf{e}}}$$

with the residuals are computed as: $\hat{\tilde{\mathbf{e}}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{b}}$. The full $mp \times mp$ covariance matrix for the estimated regression coefficients, as in eqn. (2) of Section 2.3, can then easily be defined.

In order to implement MV GLS regression, a two-stage regression approach was used:

1. *Stage 1. Ordinary Least Squares*: Carry out the MV OLS regression with its assumed error structure. The residuals are then subjected to the usual diagnostics to assess normality, independence and homogeneity of variance and appropriate transforms applied, if needed. The spatial correlation of the residuals is then assessed and modelled using variogram analysis which fits parametric forms to the sample spatial auto-correlation (see Section 3.2 for details). This yields the common spatial correlation matrix $\mathbf{V}$.

2. *Stage 2. Generalized Least Squares*: Using the spatial correlation matrix $\mathbf{V}$ of the residuals from Stage 1, obtain the square root matrix $\mathbf{K}$ and transform the response, $\mathbf{Y}$ and predictors, $\mathbf{X}$, according to the above. Then carry out the MV GLS regression on the transformed variables, as above. The transformed residuals, $\hat{\tilde{\mathbf{e}}}$, are then assessed to be sure they meet regression assumptions, in particular, that they have no discernible spatial correlation.

Further iterations on this procedure are possible, but were found unnecessary for this study.

Literature Cited

Johnson RA, Wichern DW. 2001. Applied Multivariate Statistical Analysis. Prentice-Hall. 767pp.