Spatial Autoregressive
Prepared by Rebecca E. Curtis

Definition: The spatial autoregressive model (SAR) is a generalization of the linear regression model used to account for spatial autocorrelation. The model yields better classification and prediction accuracy for many spatial datasets exhibiting strong spatial autocorrelation.

There are two families of SAR model solutions, one based on Maximum Likelihood Theory and the second based on Bayesian Statistics. Maximum Likelihood theory is used to make inferences about parameters of the underlying probability distribution of a given data set. Maximum Likelihood Theory-based SAR model solutions can be classified into exact and approximate solutions, based on how they compute the least-squares (SSE) term of the SAR solution procedure. Bayesian Statistics are a logical basis for discriminating between conflicting hypotheses. They use an estimate of the degree of belief in a hypothesis before the advent of some evidence to give a numerical value to the degree of belief in the hypothesis after the advent of the evidence. Because Bayesian Statistics rely on subjective degrees of belief, however, they are not able to provide a completely objective account of induction.

The SAR model, also known in the literature as spatial lag model or mixed regressive model, is an extension of the linear regression model and is given in the following equation:

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Here $\boldsymbol{\rho}$ is the spatial autocorrelation parameter, $\mathbf{y}$ is an *n*-by-1 vector of observations on the dependent variable, $\mathbf{x}$ is an *n*-by-*k* matrix of observations on the explanatory variable, $\mathbf{W}$ is the *n*-by-*n* neighborhood matrix that accounts for the spatial relationships (dependencies) among the spatial data, $\boldsymbol{\beta}$ is a *k*-by-1 vector of regression coefficients, and $\boldsymbol{\epsilon}$ is an *n*-by-1 vector of unobservable error. The *spatial autocorrelation* term $\boldsymbol{\rho}\mathbf{W}\mathbf{y}$ is added to the linear regression model in order to model the strength of the spatial dependencies among the elements of the dependent variable, $\mathbf{y}$. One can use Moran's I index in order to see whether there is significant spatial dependency in the given dataset.

References/Sources:

Cressie, N. 1993. *Statistics for Spatial Data*. New York: Wiley.

Kazar, B.M., S. Shekhar, D. Boley, D.J. Lilja, R.K. Pace, and J.P. LeSage. 2005. Parameter Estimation for the Spatial Autoregression Model: A Rigorous Approach. (Draft) *IEEE Transactions on Knowledge and Data Engineering.* http://www.arctic.umn.edu/papers/kazar_sar_05.pdf

Kazar, B., S. Shekhar, D. Lilja, R. Vatsavai, and R. Pace. 2004. Comparing exact and approximate spatial auto-regression model solutions for spatial data analysis. *Third International Conference on Geographic Information Science (GIScience2004).*

LeSage, J. 2000. Solving large-scale spatial autoregressive models. *Second Workshop on Mining Scientific Datasets.*

Li, B. 1996. Implementing spatial statistics on parallel computers. *Practical Handbook of Spatial Statistics*. CRC Press, pages 107–148.

Martin, R. 1993. Approximations to the determinant term in gaussian maximum likelihood estimation of some spatial models. *Statistical Theory Models*, 22(1):189–205.

Pace, R., and J. LeSage. 2002. Semiparametric maximum likelihood estimates of spatial dependence. *Geographical Analysis*, 34(1):76–90.

Pace, R., and J. LeSage. 2003. Spatial auto-regressive local estimation (sale). *Spatial Statistics and Spatial Econometrics, ed. by Art Getis*.