Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 226–259
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

A spatial model for multivariate lattice data

Stephan R. Sain[a], Noel Cressie[b,*]

[a]Department of Mathematical Sciences, University of Colorado at Denver and Health Sciences Center, P.O. Box 173364, Denver, CO 80217-3364, USA
[b]Department of Statistics, The Ohio State University, Columbus, OH 43210-1247, USA

Accepted 5 September 2006
Available online 31 January 2007

Abstract

In this article, we develop Markov random field models for multivariate lattice data. Specific attention is given to building models that incorporate general forms of the spatial correlations and cross-correlations between variables at different sites. The methodology is applied to a problem in environmental equity. Using a Bayesian hierarchical model that is multivariate in form, we examine the racial distribution of residents of southern Louisiana in relation to the location of sites listed with the U.S. Environmental Protection Agency's Toxic Release Inventory.

JEL classification: C11; C15

Keywords: Markov random field; Conditional autoregressive model; Bayesian hierarchical model; Environmental equity; Environmental statistics

1. Introduction

Many spatial problems, particularly those concerning environmental investigations, are inherently multivariate, in that more than one variable is typically measured at each spatial location. Multivariate spatial databases are becoming much more prevalent with the advent of geographic information systems (GIS) that allow users to display many different spatial data layers.

*Corresponding author. Tel.: +1 614 292 5194; fax: +1 614 292 2096.
E-mail addresses: steve.sain@cudenver.edu (S.R. Sain), ncressie@stat.ohio-state.edu (N. Cressie).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.09.010

Formally, consider a spatial location $\mathbf{s}$ and a p-dimensional random variable $\mathbf{Y}(\mathbf{s})$ associated with each location. Letting $\mathbf{s}$ vary over an index set $D \subset \Re^d$ generates a multivariate random field $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\}$. For geostatistical data, D is a given subset of $\Re^d$ and $\mathbf{s}$ is assumed to vary continuously throughout D. For lattice data, D is assumed to be a given finite or countable collection of points. Lattices may be either regular, as on a grid, or irregular, such as zip codes, census divisions (counties, tracts, block groups, or blocks), police precincts, game-management units, etc.

Models for multivariate geostatistical data (i.e., data with continuous spatial index) have been extensively explored (e.g., Wackernagel, 1998; Royle and Berliner, 1999; Ver Hoef and Cressie, 1993; Ver Hoef et al., 2004; Gelfand et al., 2004), while models for multivariate lattice data (i.e., data with a countable index set) have received relatively less attention. Mardia (1988) introduced a multivariate Markov random field (MRF) model for image processing, and more recently Billheimer et al. (1997), Kim et al. (2001), Pettitt et al. (2002), Carlin and Banerjee (2003), Gelfand and Vounatsou (2003), and Jin et al. (2005), have explored these multivariate MRF models and their role in Bayesian hierarchical modeling.

An integral feature of MRF models involves the specification of neighborhoods. For each site or lattice point $\mathbf{s}_i$, a neighborhood is a collection of sites that are spatially close. Neighboring sites can be defined, for example, as two sites separated by a fixed distance or as two sites that share a common boundary. Formally, define $\{N_i\}$ as the collection of neighborhoods, where each N_i is a set of indices representing the neighbors of $\mathbf{s}_i$. Through this specification of neighborhoods, a MRF model of the spatial-dependence structure in lattice data can be constructed.

We consider a class of hierarchical models for lattices, either regular or irregular, with n locations and (potentially) $p>1$ measurements at each location. Letting $\mathbf{Y}(\mathbf{s}_i) = (Y_1(\mathbf{s}_i), \ldots, Y_p(\mathbf{s}_i))' \equiv (Y_{i1}, \ldots, Y_{ip})'$, where Y_{ik} denotes the kth observation made at the ith lattice point, the data model is generally written as

$$Y_{ik}|\theta_{ik} \sim f(y|\theta_{ik}), \quad i = 1, \ldots, n; \; k = 1, \ldots, p.$$

Let $\boldsymbol{\theta}$ denote the $n \times p$ matrix of the process parameters and $\boldsymbol{\theta}^v \equiv \mathrm{vec}(\boldsymbol{\theta}')$, which is a $np \times 1$ vector obtained by stacking the columns of $\boldsymbol{\theta}'$. Then, the process model for $\boldsymbol{\theta}$ (or perhaps some suitable transformation of $\boldsymbol{\theta}$) follows a multivariate normal distribution given by

$$\boldsymbol{\theta}^v \sim \mathrm{N}_{np}(\boldsymbol{\mu}^v, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}^v \equiv \mathrm{vec}(\boldsymbol{\mu}')$ and $\boldsymbol{\mu}$ is the $n \times p$ matrix whose (i, k)th element is $\mu_{ik} = \mathrm{E}[\theta_{ik}]$. The large-scale dependence in $\boldsymbol{\theta}$ is captured in the mean $\boldsymbol{\mu}$, while the small-scale, spatial dependence is captured in the covariance matrix $\boldsymbol{\Sigma}$.

The covariance matrix $\boldsymbol{\Sigma}$ is determined by the neighborhood structure of the process on the lattice. Previous efforts at defining the nature of this covariance matrix have used simplified forms that, while computationally efficient, can unduly constrain the covariances; see references at the end of Section 3.2.1. These models, in general, restrict the degree of spatial dependence for different variables as well as force the dependence between different variables at different locations to be symmetric. Royle and Berliner (1999) and Ver Hoef et al. (2004) suggest flexible multivariate models for geostatistical data that include asymmetric spatial dependencies. Jin et al. (2005) propose a conditional model for multivariate lattice data but do not explicitly model the cross-dependencies. In this article, we propose a more general and flexible model for multivariate lattice data that does