ESTIMATING PREDICTION ERROR IN A SPATIAL REGRESSION MODEL

By

Nesan Chelliah*

Abstract

In this paper we consider a spatial linear regression model with random effect. Using a new estimator for the covariance matrix we have derived an expression for computing the expected value of the estimated variance of the prediction error.

Key Words: Estimated generalized least squares estimator, spatial data, Best linear unbiased estimator.

*** Address for correspondence: 11 Bolan Court, Crestmead, Brisbane, Qld 4132, Australia.**

e-mail: N.Chelliah@mailbox.gu.edu.au

1. INTRODUCTION

Air pollutant in a city often influenced by many sources in the area, such as atmospheric variables (wind speed and direction, temperature and atmospheric stability) , chemical reaction in the atmosphere , interaction with physical surface or biological system and other phenomena. Many of these process share certain traits in common and it is possible to model them using a linear spatial regression model.

Spatial modeling has increasingly been used in environmental data analysis , see for example Cressie (1994) for the modeling of chemical data and Donnelly et al (1994) for the study of air pollution and respiratory illness in children. Since environmental data often spatial over regions where the spatial index is continuous, geostatistics is having an increasingly important role to play in environmental modeling efforts, see for example Journel (1984,1988). Many of the geostatistical methods are now being used in environmental modeling.

In spatial models it is a common practice to use estimated covariance parameters in estimation. The effect of using estimated spatial covariance parameters in the estimation of regression parameters and standard errors has been discussed in Cressie (1991,pp22-24) . In a slightly different context Zimmerman and Cressie (1992) , Watkins and Mardia (1992) have delt with the problem of prediction using estimated covariance parameters.

In this paper we consider a new estimator for the covariance matrix of the dependent variable. This new estimator has recently been studied by Chelliah (1998), Anh and Chelliah (1998).

2. THE MODEL

Spatial data can be considered as a realization of a stochastic process y(u), $u \in D \subset \Re^d$ **here u is a location in D . Most often d the dimension of the space is 1,2 or 3. We assume for any u , E[y(u)] and variance of y(u) exists. Let the observations taken at location** $u_1, u_2 .....u_n$ **be** $y(u_1).....y(u_n)$. **We consider the model**

(2.1) $$Y = X\beta + v + \varepsilon$$

where Y is Nx1 observed vector , X a known Nxk design matrix and v, ε **are mutually independent vector random variables . We assume** ε **is normal with mean zero and covariance** $\sigma^2 I_N$**, and** v **is normal with mean zero and covariance matrix** Ω **. It follows that Cov(Y)=V=** $\Omega + \sigma^2 I_N$ **.**

From Christensen et al (1992) the Best Linear Unbiased Estimator of $Z = X\beta + v$ **is**

(2.2) $$\hat{Z} = X\hat{\beta} + \Omega HY$$

here $\hat{\beta} = AX'V^{-1}Y, A = Var(\hat{\beta}) = (X'V^{-1}X)^{-1}, H = V^{-1} - V^{-1}XAX'V^{-1}.$ **It follows immediately that the prediction error is**

(2.3) $$\hat{Z} - Z = (I_N - \sigma^2 H)\varepsilon + \sigma^2 Hv$$

Then from the mutual independence of the error term we have

(2.4) $$Var(\hat{Z}-Z)=\sigma^2 I_N-\sigma^4 H$$

3. PREDICTION ERROR WITH ESTIMATED COVARIANCE

Recently in Chelliah (1998) a new estimator for Ω has been proposed. The estimator is given by

(3.1) $$\hat{\Omega}=(Y-X\overline{\beta}_{LS})(Y-X\overline{\beta}_{LS})'$$

where $\overline{\beta}_{LS}=1/n\sum(X_i' X_i)^{-1}X_i' Y_i$ **. For a large sample properties of this estimator (3.1) see also Anh and Chelliah (1998). When σ^2 is known we estimate the covariance of y by substituting (3.1) .**

(3.2) $$\hat{V}=\hat{\Omega}+\sigma^2 I_N$$

We define
$\hat{Z}_E=X\hat{\beta}_E+\hat{\Omega}\hat{H}Y, \hat{\beta}_E=(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y, \hat{A}=(X'\hat{V}^{-1}X)^{-1}, \hat{H}=\hat{V}^{-1}-\hat{V}^{-1}X\hat{A}X'\hat{V}^{-1}.$
Hence Var($\hat{Z}_E-Z)=Var(\hat{Z}-Z)+Var(\hat{Z}_E-\hat{Z})$ **. It is easy to show that the covariance term is equal to zero, see for example Zimmerman and Cressie (1992). We now state and prove our main results.**

Theorem

Let $m_1 = \mathrm{var}(\hat{Z} - Z), m_2 = \mathrm{var}(\hat{Z}_E - Z)$ **be the variances as defined before. If** $\hat{m}_1$ **denotes the estimator of** m_1 **, obtained by replacing** Ω **by (3.1) then we have**

$E[\hat{m}_1] = 2m_1 - m_2$ **.**

Proof:

$$\hat{Z}_E - \hat{Z} = \sigma^2 (Q - \hat{Q})Y, Q = V^{-1}(I - P), \hat{Q} = \hat{V}^{-1}(I - \hat{P}),\ for$$
$$P = X(X'V^{-1}X)^{-1}X'V^{-1}, \hat{P} = X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}, here$$
$$Var(\hat{Z}_E - \hat{Z}) = E[\mathrm{var}((\hat{Z}_E - \hat{Z})) \mid \hat{V}] = E[\sigma^2 (Q - \hat{Q})V(Q - \hat{Q})').$$

The following identities are easy to show:

QVQ=Q , $\hat{Q}VQ = \hat{Q}, \hat{Q}\hat{V}\hat{Q} = \hat{Q}$ **, it then follows from these identities that**

E[($Q - \hat{Q})V(Q - \hat{Q})] = E(\hat{Q}) - Q$ **, hence we have**

E($\hat{m}_1) = 2m_1 - m_2$

The above expression can be easily used to compute the expected value of the estimated variance.

REFERENCES

Anh,V.V. and Chelliah,N.(1998). Estimated generalized least squares estimator in random coefficient model. Scand. J. of Statist. 25, 1-16

Chelliah,N.(1998). A new covariance estimator in random coefficient regression model. Sankhya,B (to appear).

Christensen,R., Johnson,W., and Person,L.M.(1992). Prediction diagonistic for spatial linear models. Biometrika, 79, 583-91.

Cressie,N.(1991). Statistics for spatial data. JohnWiley and Sons, New York.

Cressie,N.(1994). Spatial chemostatistics, Chapter 6, Hanbook on environmental statistics, assessment and forecasting. Lews Publication.

Donnelly,C.A.,Ware,J.H. and Laird,N.M.(1994). Regression analysis of spatially correlated data. The Kanawha county health study. Hanbook of statistics, vol 12, 643-660.

Journal,A.G.(1984). New ways of assessing spatial distribution of pollutants. In environmental sampling for hazardous wastes,American chemical society, Washington DC, 109-118.

Journel,A.G.(1988). Nonparametric geostatistics for risk and additional sampling assessment. In principal of environmental sampling,American chemical society, Washington DC, 45-72.

Watkins,A.J. and Mardia,K.V.(1992). Maximum likelihood estimation and prediction mean squares error in the spatial linear model. J. Appl. Statist. 19, 49-59.

Zimmerman,D.L. and Cressie,N.(1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. Ann.Inst.Statist. Math., 44, 27-43.