A Prediction Approach to Agricultural Land Use Estimation Using a High-Resolution Spatial Model of Land Use Dynamics

Ambrosio L.
Marín C.
Iglesias L.

Universidad Politécnica de Madrid. Spain

Abstract

The prediction approach for finite populations using (super population) linear models is now well - established. Nevertheless, for count and binary data non linear models are more adequate. In this study a generalized linear mixed model (multinomial) of land use decisions is proposed to estimate agricultural land uses, in a prediction approach framework. The spatiotemporal correlation structure is assessed using semivariograms models and a panel data set on land uses and on auxiliary variables observed in a spatial sample. The model is fitted using a generalized estimating equation approach. Parameter estimators and predictors, as well as their standard error estimators, are consistent, even if the working spatiotemporal correlation structure is not the true one. They are biased in small samples and the degree of bias increases substantially if spatiotemporal correlation is ignored.

The model

Let $\underline{Z}\left(a_{s_i t}\right)=\left\{Z_k\left(a_{s_i t}\right);\ k=1,2,\cdots,K\right\}$ be the acreages on K land use categories within the sampling unit $a_{s_i t}$, where s_i is a vector of coordinates which locates $a_{s_i t}$ within the studied region. The working model is $\eta_k\left(a_{s_i t}\right)=\log(\mu_k\left(a_{s_i t}\right))=\underline{x}_k\left(a_{s_i t}\right)\underline{\alpha}_k+\underline{Z}_k\left(a_{s_i t}\right)\underline{\phi}_k$ where $\mu_k\left(a_{s_i t}\right)=E\left(Z_k\left(a_{s_i t}\right)\middle|\underline{x}_k\left(a_{s_i t}\right),\underline{Z}_k\left(a_{s_i t}\right)\right)$ and $\underline{x}_k\left(a_{s_i t}\right)$ is a row vector of $(1\times p)$ explanatory variables of known values, $\underline{\alpha}_k$ is a column vector of $(p\times 1)$ unknown parameters, $\underline{Z}_k\left(a_{s_i t}\right)$ is the $(1\times K(T-1))$ vector of the past observations $\left\{\underline{Z}\left(a_{s_i,1}\right),\underline{Z}\left(a_{s_i,2}\right),\cdots,\underline{Z}\left(a_{s_i,T-1}\right)\right\}$ until time T and $\underline{\phi}_k$ is a column vector of $(K(T-1)\times 1)$ unknown parameters. The covariance matrix is $\underline{V}_k=\left[\upsilon^{1/2}(\underline{\mu}_k)\right]R(\underline{\theta}_k)\left[\upsilon^{1/2}(\underline{\mu}_k)\right]$ where $\upsilon(\underline{\mu}_k)=diag\left\{\upsilon(\mu_k\left(a_{s_i t}\right))\right\}$ and $\upsilon(\mu_k\left(a_{s_i t}\right))$ are the variance functions. $R(\underline{\theta}_k)$ is the correlation matrix, where $\underline{\theta}_k$ is the parameter's vector of the spatiotemporal correlation structure: the process is assumed to be second order stationary and isotropic and a product-sum spatiotemporal function is considered as variogram [Iaco *et al.* (2001)]. The model specified for the whole NTK sampling data is $\underline{\eta}=g(\underline{\mu})=\underline{\tilde{X}}\underline{\beta}$ and the covariance matrix is $\underline{V}=diag\left\{\underline{V}_k\right\}$. The estimator $\hat{\underline{\beta}}$ of $\underline{\beta}$ is obtained by iteratively reweighted least squares (McCullagh and Nelder, 1989) using an estimating equations approach (Zeger and Liang, 1986; Liang and Zeger, 1986).

The predictor

The predictor of $\underline{Z}\left(a_{s_jT}\right)=\left\{Z_k\left(a_{s_jT}\right);\ k=1,2,\cdots,K\right\}$ where $Z_k\left(a_{s_jT}\right)$ is the k^{th} land use acreage observed in the sampling unit $a_{s_j,T}$ not included in the sample is [Valliant (1985), Gotway and Stroup, 1997)] $\hat{\underline{Z}}\left(a_{s_jT}\right)=\hat{\underline{\mu}}\left(a_{s_jT}\right)+\underline{C}_{s_jT}^{T}\underline{V}^{-1}(\underline{Z}-\hat{\underline{\mu}})$ where elements in $\underline{C}_{s_jT}^{T}$ are as $C_{ks_jT,s_it}=Cov\left(Z_k\left(a_{s_jT}\right),Z_k\left(a_{s_it}\right)\right)=\upsilon^{1/2}\left(\mu_{ks_jT}\right)\rho(dist\left(s_j,s_i\right),\left|t-T\right|;\underline{\theta}_k)\upsilon^{1/2}\left(\mu_{s_itk}\right)$ where $\rho(.)$ is the correlogram function. The performance of $\hat{Z}_k\left(a_{s_jT}\right)$ and of $V\left(\hat{Z}_k\left(a_{s_jT}\right)-Z_k\left(a_{s_jT}\right)\right)$ was evaluated by Monte Carlo. Results for one hundred repetitions of a panel data set with $n=110$ and $T=7$ are shown in the following table.

Predictor performance

Quantiles	Predictor Bias (%)	Standard error bias (%)		Coefficient of Variation (%)
		Taking into account spatiotemporal correlation	Ignoring spatiotemporal correlation	
$Q_1(25\%)$	-5.06	-24.87	-47.34	4.01
$Median(50\%)$	-1.49	-14.92	-33.79	10.51
$Q_3(75\%)$	1.95	-4.67	-23.66	16.13

REFERENCES

De Iaco, S., Myers, D.E. and Posa, D. (2001). Space-time analysis using a general product-sum model. *Statistics & Probability Letters* 52: 21-28.

Gotway, C.A. and Stroup, W.W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological and Environmental Statistics* 2: 157-178.

Liang, K.Y. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika.* 73: 13-22.

McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. Chapman and Hall: London.

Valliant, R. (1985). Nonlinear prediction theory and the estimation of a proportion in a finite population. *Journal of the American Statistical Association* 80: 631-641.

Zeger, S.L. and Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42: 121-130.